The Dissertation Committee for Ajay Gopinath
certifies that this is the approved version of the following dissertation:

Feature Modeling and Tomographic Reconstruction of Electron Microscopy Images

Committee:

Alan C. Bovik, Supervisor

David Ress, Co-Supervisor

Guoliang Xu

John Pearce

Joydeep Ghosh

Feature Modeling and Tomographic Reconstruction of Electron Microscopy Images

by

Ajay Gopinath, B.E., M.S

DISSERTATION

Presented to the Faculty of the Graduate School of
The University of Texas at Austin
in Partial Fulfillment
of the Requirements
for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN
May 2012

Dedicated to Mitali and Ishan.

Acknowledgments

I would like to express my deepest gratitude to Prof. Al Bovik for his unwavering support and sound advice. He continues to provide inspiration and it has been an honor to be his student. I would also like to thank Prof. David Ress for all the great discussions we had and his continued support. I would like to thank Prof. Guoliang Xu for all his help. I was fortunate to work with him at the Chinese Academy of Sciences, which was a wonderful experience. I would like to thank my committee members for their valuable inputs. I am grateful to all my colleagues in the CVC and LIVE lab, especially Radhakrishna Bettadapura (RK), Dr. Qin Zhang, Maysam Moussalem, Jose Rivera, Suzanne Bailey, Anush Murthy and Gautam Muralidhar for all the wonderful discussions that made graduate school enriching. I wish to thank the department of Electrical and Computer Engineering, especially the very patient Melanie Gulick who helped out with the University rules and requirements.

I would like to thank my parents and sister who have been a constant source of support and inspiration throughout my life, for which I am deeply indebted to them. Finally, I would like to thank my wife Mitali, for her wonderful support throughout my PhD and our son Ishan who is a constant source of joy in our lives.

Feature Modeling and Tomographic Reconstruction of Electron Microscopy Images

Ajay Gopinath, Ph.D.
The University of Texas at Austin, 2012

Supervisors: Alan C. Bovik
David Ress

This work introduces a combination of image processing and analysis methods that perform feature extraction, shape analysis and tomographic reconstruction of Electron Microscopy images. These have been implemented on images of the AIDS virus interacting with neutralizing molecules. The AIDS virus spike is the primary target of drug design as it is directly involved in infecting host cells. First, a fully automated technique is introduced that can extract sub-volumes of the AIDS virus spike and be used to build a statistical model without the need for any user supervision. Such an automatic feature extraction method can significantly enhance the overall process of shape analysis of the AIDS virus spike imaged through the electron microscope. Accurate models of the virus spike will help in the development of better drug design strategies.

Secondly, a tomographic reconstruction method implemented using a shape based regularization technique is introduced. Spatial models of known features in the structure being reconstructed are integrated into the reconstruction process as regularizers. This regularization scheme is driven locally through shape information obtained from segmentation and compared with a known spatial model. This method shows reduced blurring, and an improvement in the resolution of the reconstructed volume was also measured. It performs better than popular current techniques and can be extended to other tomographic modalities. Improved Electron Tomography reconstructions will provide better structure elucidation and improved feature visualization, which can aid in solving key biological issues.

Table of Contents

List of Figures

Chapter 1

Introduction

Electron microscopy (EM) is the preferred tool for structural biologists to visualize three-dimensional structures of molecular and cellular complexes *in-situ*. Electron tomography (ET) involves acquiring planar Transmission Electron Microscope (TEM) images of the biological sample from different projection angles (tilt series). Simulating this procedure is called the *forward problem*, and reconstructing a 3D image (or map) from such data using the principles of tomography is called the *inverse problem*. ET is currently the *only* approach that allows one to reconstruct the 3D structure of individual biological complexes in their native state.

The reconstruction from ET data must overcome several challenges. First, EM images suffer from limited contrast to noise ratio. A second major challenge is that the angle of rotation (tilt angle) cannot exceed ±70° (with respect to the horizontal plane) because the beam's path-length through the sample and its supporting structure becomes inappropriate to form projections at higher tilt angles. Hence projection images are available only for a limited range of tilt angles. In Fourier space, the missing tilt images manifests themselves as a missing wedge (Fig. 1.2), leading to severe instability in the reconstruction problem. Due to the missing data, the inverse problem is ill posed and there is no unique solution to the reconstruction. Thus, generation and visualization of models from ET data is a major challenge.

One of the major objectives of this thesis is to develop mathematical and computational tools to solve the *inverse problem* in ET through a shape regularization method derived from a Bayesian formulation. Likelihood probability distributions that accurately model the forward problem in ET as well as the main noise characteristic in the measured data were formulated. Structural shape information about the specimen being imaged was included into a shape based regularization by means of an energy functional. Reconstruction was performed by calculating the maximum *a posteriori* (MAP) estimate through a modified Expectation Maximization (EM) optimization scheme.

1.1 The AIDS Virus

The AIDS virion is roughly spherical in shape, with an inner capsid region that encloses its genome and an outer proteinaceous envelope on which several protruding entities called *spikes* are distributed (Fig. 1.1). Each spike is roughly mushroom shaped with a tapering stem that is attached to the virus envelope. The head of the mushroom shaped structure has a trimeric protein known as gp120, each of whose monomeric subunits is arranged symmetrically around an axis passing through the center of the spike. A cylindrically shaped protein known as gp41 connects with the proteinaceous envelope. The virus particle is typically $120nm$ in diameter while the height of the spike is around

$120\mathring{A}$ with a maximum width of about $150\mathring{A}$, tapering to $35\mathring{A}$ at the junction of the envelope [25].

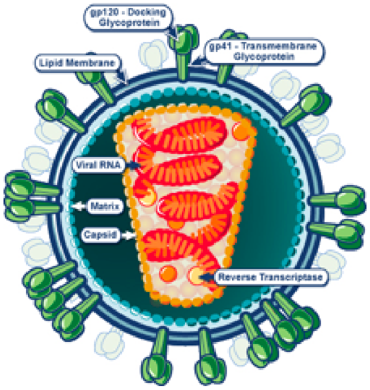


Figure 1.1: A schematic of the HIV virus shown here. The virus spikes are the gp120 and the gp41 regions that protrude out of the virus envelope. (source: www.niaid.nih.gov)

1.2 Electron Microscopy Preliminaries

When a high energy electron wave interacts with a biological specimen, it is scattered due to the electrostatic field of the specimen's constituent atoms. The output wave after passing through the specimen contains information about the electrostatic field of the specimen which is captured as image contrast. The main contrast mechanism in EM imaging of unstained biological samples is phase contrast (interference contrast), which results from the quantum superposition (interference) of the crests of the incident wave and the scattered wave. This can often be approximately described by projecting the potential of the specimen. Typically, thinner specimens with higher accelerating voltage and lighter atoms will satisfy this projection approximation [9] (see Appendix 1 for details of the physics of EM image formation). The emergent electron wave interacts with the microscope optics to generate a 2D projection image. The specimen is tilted and several such 2D projection images are acquired. The 2D projection images in EM typically suffer from a number of degrading factors. The images acquired have spatial resolution on the order of Angstroms (Å), but slight mechanical noise during tilting and acquisition can cause large misalignment in the images. These factors combined with the missing wedge problem described earlier degrade the accuracy of 3D reconstructions in ET.

The inverse problem in ET is to reconstruct a 3D image based on 2D images acquired within a tilt series. Mathematically this can roughly be stated as the problem of recovering an estimate of a signal f (a function representing the 3D image) given measured data g (the vector representing data in the tilt series) obtained from indirect observations of the signal. The precise mathematical formulation depends on whether one takes the statistical or classical approach for regularization. All regularization methods, both classical and statistical, are based on the principle of replacing the original ill-posed inverse problem with a neighboring well-posed problem. This is done by making use of a prior model and a data model.

Prior model: A prior model encodes the a priori information about the signal to be recovered or some property thereof. Commonly occurring examples are positivity and specific regularity.

Data model: A data model contains assumptions about how the data relates to the unknown signal and the measurement errors. An important component of the data model in both the classical and statistical approaches is the modeling of the uncorrupted data which is given by the forward model H. This is a deterministic model for the experiment that relates the signal to corresponding uncorrupted data in absence of noise and measurement errors.

The projection images collected are denoted by

$$I_\theta(x, y) : 0 \leq x < X,\ 0 \leq y < Y, -70^\circ \leq \theta \leq +70^\circ.$$

The size of each projection image X, Y is determined by the microscope hardware and the range of the tilt angle θ is set by the biologist and the capabilities of EM hardware. The forward model for ET is derived in [8] where it is shown that under certain conditions, one can assume that the tilt series images correspond to a simple forward projection transform (Radon transform) of the specimen. Hence, if the function f representing the 3D image is discretized and represented by a vector, typically by taking voxel values, then the forward model can be written as a projection matrix H. For noise-free data

$$Hf = g \tag{1.1}$$

where g is the data from the projection images I_θ. The data model for ET is derived in [40] and is quite complex. But the dominating contribution to the noise is from the shot noise that is Poisson distributed.

The reconstruction problem in classical regularization is formulated as follows:

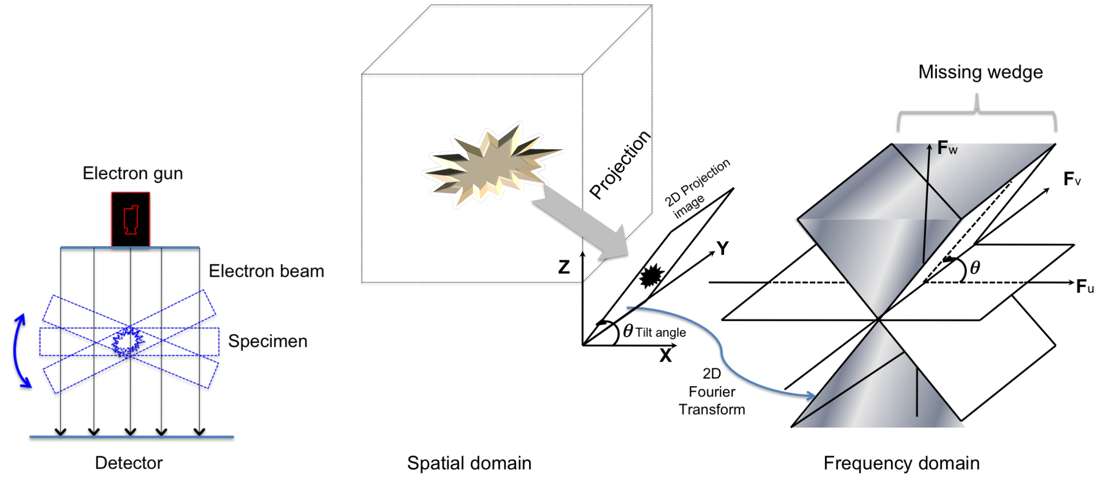
> Assume one is given measured data g, which is a corrupted version of $H(f)$ where the precise relation between g and $H(f)$ is given by the data model. Recover f from such data.

In statistical regularization, and in particular Bayesian inference, the philosophy is to recast the inverse problem in the form of a statistical quest for information. While classical regularization seeks to recover a single estimate of the unknown signal, statistical regularization is based on recovering the probability distribution of the signal. In Bayesian inference an inverse problem is formulated as follows:

> Given two random variables F and G where samples of F represent possible signals and samples of G represent possible data. A single sample of G which is the measured data g, is given. The goal is to recover the posterior probability distribution of F when the likelihood of data and a prior probability distribution of F are known.

1.3 Tomographic reconstruction methods

These are commonly subdivided into *analytical*, *iterative*, and *variational* methods.



(a) (b)

Figure 1.2: (a) The Electron Tomography (ET) image acquisition process. (b) Projection-slice theorem: The missing wedge region in Fourier space is due to limited angle projections in the spatial domain.

Analytical reconstruction methods The starting point is to consider the operator equation

$$H(f) = g$$

where the data is considered as continuum. The idea is to directly approximate H^{-1} and appropriately discretize it. Doing this in a manner that preserves stability depends on the specific properties of the operator H and the data collection geometry (the interpretation of the discretization of g). For the case when H represents a simple forward projection transform, the Filtered Back Projection (FBP) method is well known. It is based on the projection-slice theorem, which states that each projection of an image volume is equivalent to a slice through the Fourier transform of that volume sampled at the same orientation as the projection [22]. Weighted Back Projection (WBP) reconstruction is another analytic method that is popular in ET [19], [34]. It is based on the same aforementioned projection-slice theorem. Both FBP and WBP methods can be considered as special cases of mollifier methods [26].

Iterative methods Here the idea is to generate an iterative sequence that converges to a least-squares solution of $H(f) = g$ [17]. The iterates are stopped early for stability, so the iteration index acts as a regularization parameter. The most popular iterative method in ET is the Kaczmarz method, also called Algebraic Reconstruction Technique (ART) [22]. In this technique each tomogram image $\boldsymbol{g}_i$ is treated like a hyperplane in a N dimensional space.

$$\boldsymbol{g}_i = \boldsymbol{h}_i \cdot \boldsymbol{f}, \tag{1.2}$$

where $\boldsymbol{h}_i = [h_{i1}, h_{i2}, ..., h_{iN}]^T$ is the ith column of the matrix representing the forward operator H and $\boldsymbol{f} = [f_1, f_2, ..., f_N]^T$.

A solution to the entire equation system for all tomograms $\boldsymbol{g}$ is a point in the N dimensional space where all the hyperplanes intersect [22]. The update equation for ART is given by:

$$\boldsymbol{f}^{(i+1)} := \boldsymbol{f}^{(i)} - \lambda \frac{\boldsymbol{h}^{(i)} \cdot \boldsymbol{f}^{(i)} - \boldsymbol{g}_{i+1}}{||\boldsymbol{h}_{i+1}||^2} \boldsymbol{h}_{i+1}, \tag{1.3}$$

where $\boldsymbol{f}^{(i)} = [f_1^{(i)}, f_2^{(i)}, .., f_N^{(i)}]^T$.

Variational methods The idea here is to reformulate the reconstruction problem as an optimization problem:

$$\arg\min_{f} \lambda S(f) + D(H(f), g) \tag{1.4}$$

In the above, the functional $f \mapsto S(f)$ is the regularization functional that encodes a priori information about the true unknown signal, D is the data discrepancy functional, and $\lambda > 0$ is the regularization functional.

A common approach in many imaging modalities is to select $S(f) := \|\nabla f\|_2$, i.e. as the L^2-norm of the gradient magnitude of f (Dirichlet energy). This results in smoothed images. Total Variation (TV) is a popular method that attempts to preserve boundaries/edges. In TV, $S(f) := \|\nabla f\|_1$, i.e. the L^1-semi norm of the gradient magnitude of f. The precise form of the data discrepancy functional depends on the data likelihood, e.g. it becomes a (weighted) L^2-distance in data space if noise in data is additive Gaussian. Refer to [42] for a detailed treatment of variational methods in the inverse problem context.

TV and related methods are applied in [39] to the reconstruction problem in ET with promising results. Higher order regularizers have also been applied to ET. A recent method introduced by Xu et al. [49] [24] reconstructs f by solving (1.4) with

$$S(f) := \lambda \int u\big(\sigma(f)(x), \kappa(f)(x)\big)\big|\nabla f(x)\big| dx,$$

and

$$D\big(H(f), g\big) := \sum_{i} \int \big(h_i f(x) - g_i\big)^2 dx$$

where $\sigma(f)(x)$ and $\kappa(f)(x)$ are the mean and Gaussian curvatures of the level-set surface of f at a point x, and u is a user specified function. The first term is the fidelity term while the second term is a geometric regularizer. In [49], the solution of (1.4) with S and D as above is obtained by solving an L^2-gradient flow (L2GF) using a calculus of variations optimization method.

1.4 Outline

In chapter 2, a composite segmentation method is introduced, which combines segmentation outputs from different reconstruction techniques using a voting method. Geometry processing methods are outlined that attempt to build a smooth surface reconstruction of the AIDS virus spike region. Some skeletonization results are also shown. In chapter 3, a fully automated method for detecting spikes and combining the spike sub-volumes to generate a statistical shape model of the spike is introduced. In chapter 4 a new tomographic reconstruction method is introduced. This method incorporates prior shape models of known structures in the reconstruction process.

Chapter 2

Modeling From Electron Microscopy Images

In this chapter, various segmentation and current state of the art geometry processing techniques are outlined, with the overall goal being improved structure elucidation of the virus envelope and spike regions.

2.1 Composite segmentation

A novel technique for composite segmentation is presented where reconstructed volumes from four different methods and their segmented maps are combined. This method annotates the segmented output with a confidence level based on a voting scheme. This enhances structure visualization and provides comprehensive structure information with error and uncertainty annotations at a voxel level. This method helps in eliminating errors that would arise if only a single reconstruction method was used for segmentation. The variation seen in the segmentation from different reconstruction methods motivates further research for accurate reconstruction techniques.

Image segmentation was performed to extract the virus envelope and spikes from the background. The segmentation method described in [50] was used. Two point classes, one for the virus region and another for the background, were manually seeded. These seeds are initialized as *snakes* whose movements are directed by an external force determined by the normalized gradient vector field [50]. When the snake reaches an edge, it stops since the gradient vector will be opposing the direction of movement. A partition of the image was obtained corresponding to each snake. Figure 2.1 shows 2D slices of segmented maps from each reconstruction method. L^2GF and ART showed better segmentation of the virus spikes. There were voxels on the inside of the virus envelope that are erroneously segmented in ART. In SIRT the spikes were not segmented completely. The gap in the envelope as seen in ART also existed in SIRT. In WBP and L^2GF, there was no gap in the envelope as seen in ART and SIRT. While the spikes were over segmented in WBP, L^2GF had well segmented spikes.

A composite segmentation map was created by combining segmentations of all four reconstruction methods. If a voxel was segmented only by one reconstruction method it was assigned a low confidence label. Voxels that were segmented by exactly two reconstruction methods had a medium-low confidence level. Voxels that appear in segmentations of three reconstructions were tagged medium-high confidence. Voxels that were segmented in all the reconstruction methods were assigned the highest confidence level. The virus envelope was segmented with high confidence as seen in Figure 2.1 (Composite) while the erroneous non-envelope voxels were indicated as noise (low confidence). The green color shows high confidence voxels that were segmented by all the methods. Figure 2.2 shows a volume rendering of the 3D compos-

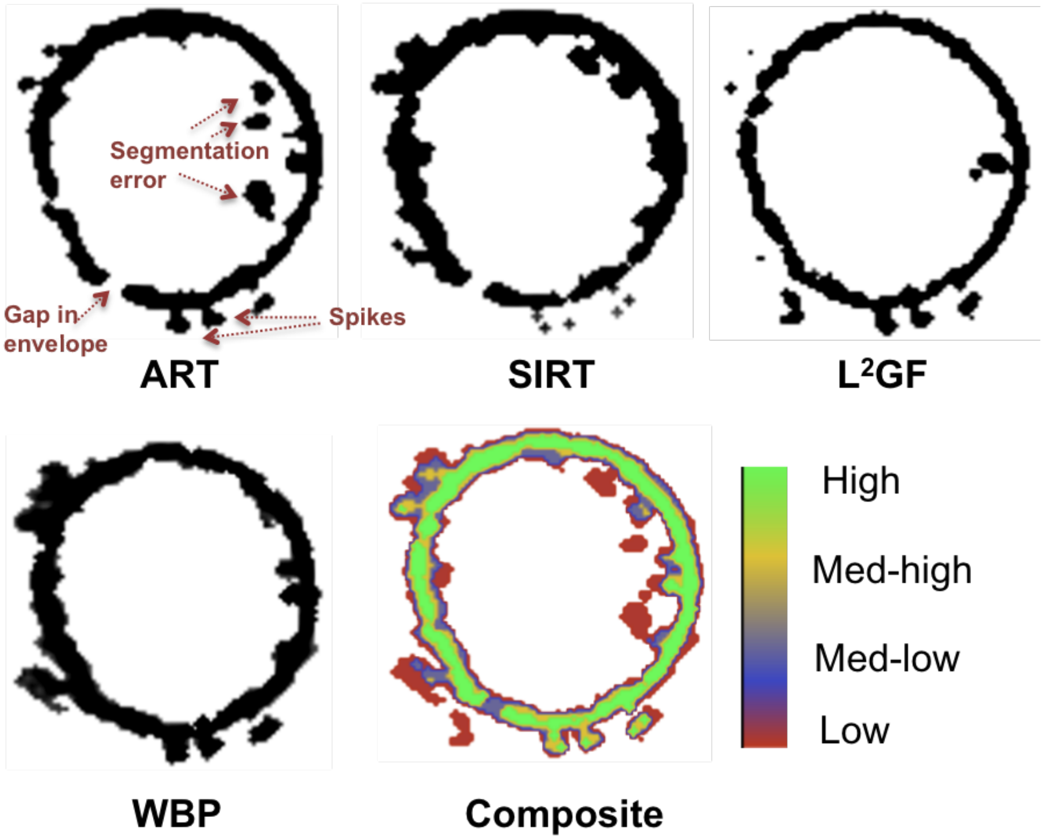
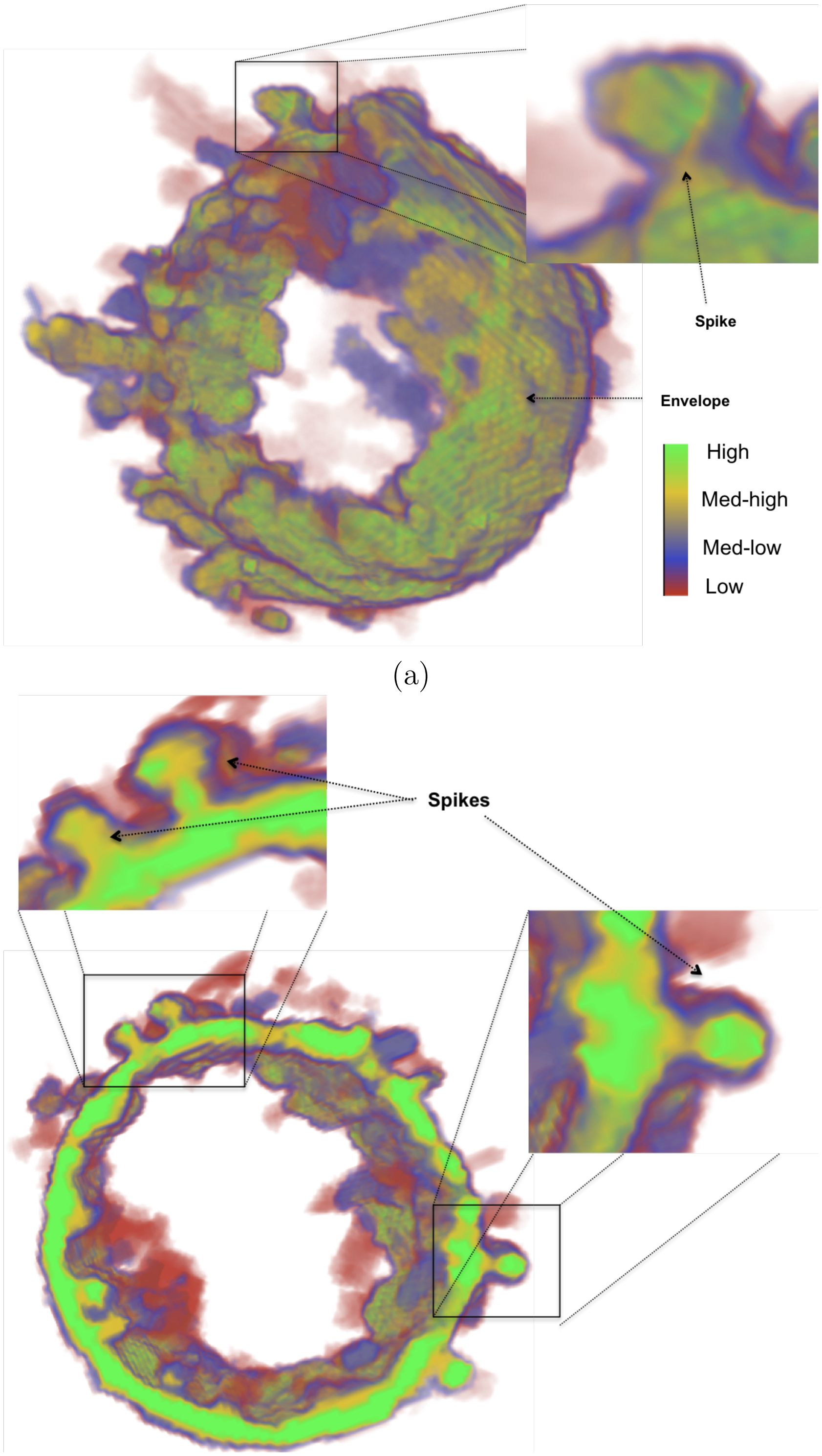


Figure 2.1: 2D slices of the segmented AIDS virus reconstructed using ART, SIRT, L^2GF and WBP. Composite Segmentation: Each voxel was assigned one of four confidence labels. A voxel segmented in all four reconstruction methods has high confidence (green), three reconstruction methods is med-high confidence (yellow), two reconstruction methods is med-low confidence (blue) and one reconstruction method is low confidence (red). Voxels that werre incorrectly segmented can be easily removed based on as low confidence values.

ite segmentation volume. This provides a 3D visualization of the segmented voxels and their confidence level. The Volume Rover software package [2] was used for visualization. The number of high confidence voxels in each segmented volume as a percentage of the total number of voxels segmented in that volume was determined. $L^2GF(45.1\%) > ART(43.26\%) > SIRT(41.16\%) >$



(a)

(b)

Figure 2.2: Volume rendering of the composite segmented AIDS virus. In (a) and (b) each voxel is tagged with a confidence level based on the number of reconstruction methods that segmented it. The spikes can be segmented with high to medium high confidence. Voxels segmented incorrectly by a reconstruction method are classified as low confidence. (b) shows a volume rendering with a clipping plane through the virus.

$WBP(39.49\%)$.The segmentation of the volume reconstructed using L^2GF has a larger number of high confidence voxels as a percentage of total voxels segmented in that volume.

2.2 Geometry Processing

In this section different geometry processing methods to reconstruct the surface of the virus and obtain a skeletonization is described. The input to this stage is a segmented volume, which can either be the high-confidence voxels segmented from the composite segmentation technique or a single segmentation output.

2.2.1 Detecting Boundary Points

The first step for surface reconstruction is to detect boundary points, which was done using a scheme similar to the segmentation step. The gradient vector flow based region growing stops at the boundary of the segmented volume, the points where the region growing stops was identified as boundary points.

2.2.2 Surface Reconstruction

A surface reconstruction was performed over the boundary point cloud using the technique described in [7]. This method provides theoretical guarantees for accurate surface reconstruction in presence of some noise model. The model allows the points to be scattered around the sampled surface with a restricted range. The algorithm works with the Delaunay/Voronoi diagrams of the input point sample. The reconstructed model is shown in figure 2.3(a).

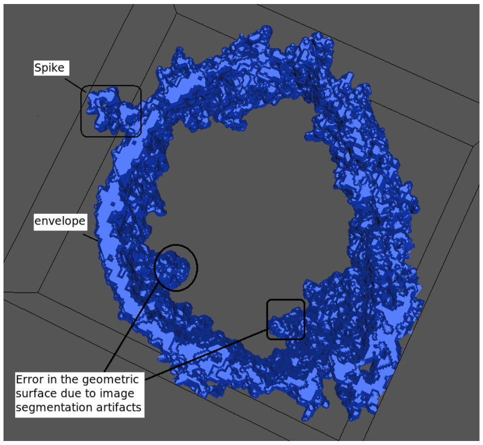
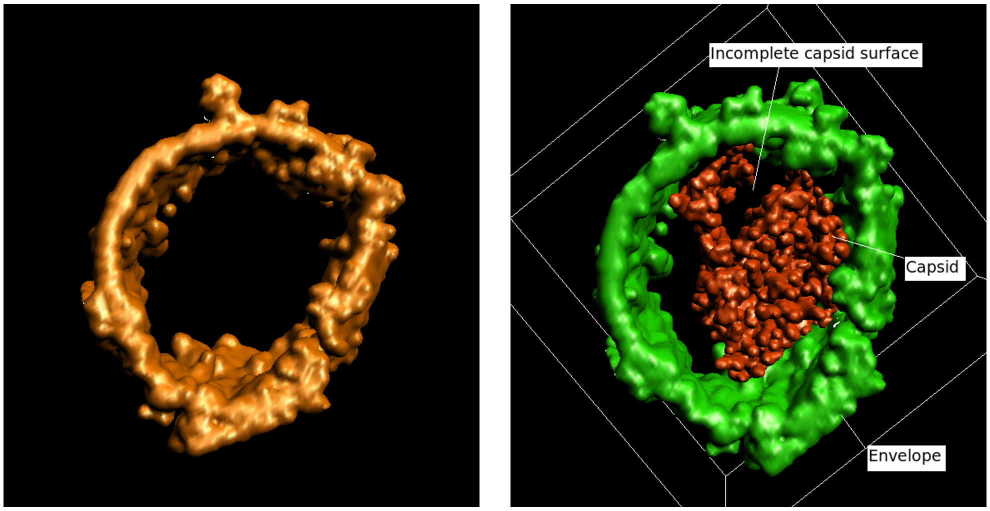


Figure 2.3: The 3D geometric surface of the segmented volume is reconstructed using the tight cocone technique described in [7]

2.2.3 Higher Order Levelset (HLS)

A higher order level set based smooth surface reconstruction was performed using the method described in [1]. The higher order comes from the cubic spline surface that is used for the surface reconstruction [1]. The topology of biological structures can be complex making surface tracking difficult. Level set surfaces obtained from classical methods require a dense grid surface to produce good quality surfaces, which increases computational cost. The method used here [1] uses tri-cubic splines as the level set function basis, which produces good quality surfaces without a large cost. The surface generated from this method is smoother than the surface reconstruction performed earlier but it does not ensure theoretical guarantees for accurate reconstruction.

To test the HLS surface reconstruction on multiple surfaces, the virus envelope and the capsid were segmented through user supervision using the techniques described above (Fig. 2.4). Two smooth surface reconstructions are generated one for the virus envelope and another for the capsid (Fig. 2.4b). Since the surface reconstruction depends on the segmentation step, errors due to segmentation translates to errors in the surface reconstruction as well.

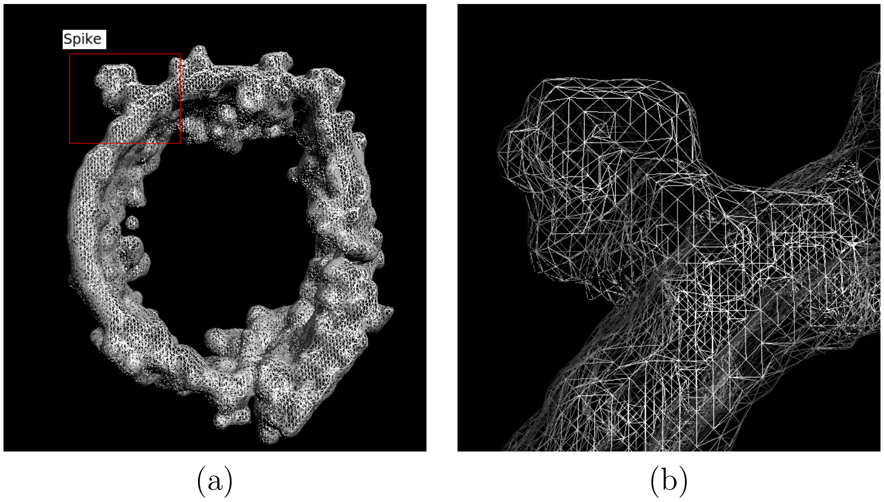


(a) (b)

Figure 2.4: (a) Isocontour surface showing the output of the HLS based smooth surface reconstruction of the virus envelope. (b) The HLS reconstruction is applied on the combined envelope+capsid data. The inner capsid shown in red appears broken due to incomplete segmentation. The error in segmentation due to noise cascades into the surface reconstruction

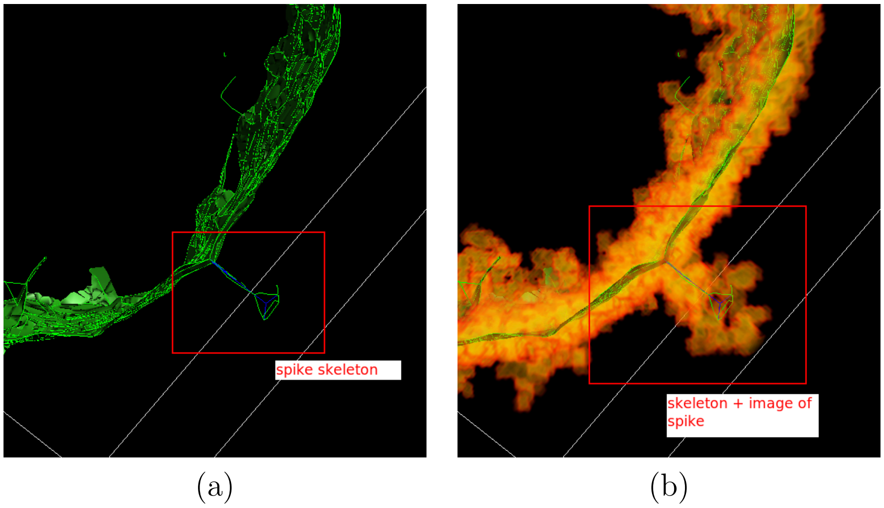
2.2.4 Mesh Regeneration

A new surface mesh is recomputed but now using the boundary points from the HLS based surface reconstruction. The HLS based surface reconstruction cannot guarantee accurate reconstruction, which the Delaunay ball based method of Dey T. et al. [7] provides. But the point cloud from the HLS step is less noisy and smoother than the boundary point cloud used previously for surface reconstruction (Section 2.2.2). Using the HLS smoothing step as preprocessing for the mesh generation, better quality surface meshes are obtained (Fig. 2.5) that are smoother and also have the theoretical guarantees (water tight surfaces with no gaps) that Dey T. et al. [7] provides.



(a) (b)

Figure 2.5: (a) Mesh generated from surface reconstruction (b) Virus spike region.



(a) (b)

Figure 2.6: (a) The virus spike after skeletonization is shown in green. The spike is attached to the skeleton of the envelope which is a green sheet like structure (b) The 3D image in orange and the skeletonization in green is combined in this image. The spike skeleton is clearly seen to be a reduced structure of the overall spike

2.2.5 1D/2D Skeletonization

Skeletonization is the process of reducing an image to a skeletal remnant that preserves the extent and connectivity of the overall feature. For biomedical images, it is useful to study the skeleton of a feature for drawing comparisons with atomic level models. A useful application is in the problem of fitting atomic information obtained from X-ray crystallography into models generated from EM images. The atomic models can be aligned with skeletal models of the features instead of aligning the atomic models with the EM image model directly.

The 1D/2D skeleton of the 3D surface is generated using the technique developed by Goswami et al. [12]. This algorithm identifies the flat and tubular regions of a three dimensional shape from its point samples by locating stable and unstable manifolds. These unstable manifolds are mapped back on to the surface resulting in tubular and flat regions. The algorithm takes as input the surface mesh reconstruction of a feature built in the previously (Section 2.2.4). Given the nature of the skeletonization algorithm, smoother surface reconstructions result in good quality skeletons. Figure 2.6 shows skeletonization of the virus spike region. The envelope is reduced to a flat sheet like region and the skeleton is reduced to an elongated linear structure.

2.3 Discussion

In this chapter various state of the art image processing and geometry processing techniques were used on EM images of the AIDS virus spike region. Since these techniques do not robustly achieve good quality models of the virus features, two new techniques are presented in the subsequent chapters. One of them is a fully automated spike detection and model generation method and the other is a new tomographic reconstruction method that attempts to address the shortcomings for current reconstruction methods.

Chapter 3

Automatic Feature Extraction and Statistical Shape Modeling

3.1 Introduction

The spike is the primary target for drug design as it allows the virus to infect the immune cells by binding and fusing with them. The precise structure and various possible states of the virus spike is of high importance for biochemists who design drugs that can neutralize the AIDS virus. Shape complementarity between the drug and the virus spike is one of the critical aspects of drug design. Currently, biochemists identify spikes and segment them through manual supervision or by semi-automated methods where a user provides the initial locations or inputs to a segmentation algorithm that extracts spike features within a user defined subvolume [25] [28]. Liu et al. [25], Zhu, et al. [54] and others have extracted several individual spikes using manual processes and performed alignment and averaging to create a single averaged spike model [25]. Spike extraction methods involving user supervision can be tedious and time consuming. The objective is to use image processing and computer vision to fully automate the process of detecting and extracting spikes without the need for any user supervision. The efficacy of this method is demonstrated by also producing a statistical shape model of the spike instead of a single average shape model as reported in the literature [25], [54]. This fully automated process could significantly enhance the overall spike analysis pipeline, providing biochemists involved in drug design the ability to process virus data in much larger numbers leading to more accurate structure elucidation. This algorithm is the first method that directly addresses the spike detection and statistical model generation in a fully automated framework.

3.1.1 Statistical Shape Models

A single template shape is not sufficient for most biological structures due to their high variability. A statistical model aims to include common variations of the structure in the model. The most common and simplest method to represent shapes is a set of points that are distributed across the structure's surface. These points are commonly referred to as landmarks, though they need not be located at salient feature points as per the common definition for anatomic landmarks [18]. Landmarks have been used to build statistical shapes of biological structures by Bookstein [4] and others. Medial axis models or skeletons have also been used to describe biological shapes. The structure is represented by centerlines and corresponding radii. Pizer et al. [32] introduced a medial model with a coarse-to-fine representation that uses a collection of points on centerlines and vectors towards the boundary. A non-uniform rational B-Splines (NURBS) method was used by Tsagaan et al. [46] to model a variety of objects, including the kidney that possess intricate features. Methods that use landmarks need to ensure that they are all located on corresponding positions across all the training samples. Obtaining the correspondence of landmarks across several 3D volumes is not trivial. Typical methods to construct a statistical shape involve extracting a mean shape and modes of variation from training samples. The first step is to align the shapes; there are several methods for aligning both rigid and non-rigid objects. The aligned data can be compactly represented using methods such as principal component analysis (PCA) that represent shapes by a linear combination of modes. In general, PCA results in global modes which influence all variables simultaneously, hence varying one model will affect the entire shape [18].

Other methods include those by Cootes and Taylor [6] that use finite element methods to calculate vibrational modes for the training data and that are used to create a model that can represent all shape instances. To increase the model flexibility, some approaches split the statistical shape into different parts, where each part varies independently. Zhao et al. [53] create a multi-partite model by using mesh partitioning, where each part of the mesh is modeled separately. Rueckert et al. [38] employ statistical deformation models (SDM) to construct anatomical models of the brain. This method is closely related to the developing field of Computational Anatomy (CA) method promoted by Grenander and Miller [15]. Computational Anatomy involves generating models from a set of anatomical images [16] [27] [15]. The idea in CA is to carry out statistical analysis directly on deformation fields that are obtained by performing non-rigid registration, without the need for segmentation and correspondence estimates. Also, instead of performing analysis on the deformation field directly, the statistical analysis is performed on the control points of the deformation fields. The advantage of these control points is in providing compact representation of the deformation field. Methods described by Rohlfing et al. [36] use repeated application of a non-rigid registration method based on B-splines to generate an average model.

To build a statistical shape model of the AIDS virus spike, spike subvolumes that are extracted and aligned automatically were used. The intensity information from all the individual detected spikes are combined. The resulting statistical model of the spike is 4D, where the fourth dimension is a probability density function assigned for that voxel. The density function is constructed at each voxel based on the samples from all the detected spikes.

3.2 Method

Volumetric images are generated from the tomographic reconstruction of single axis tilt series images taken from the range ±69° from a TEM. The images are of the Simian Immunodeficiency Virus (HIV-like retrovirus that causes AIDS in monkeys) [3]. A Maximum Likelihood reconstruction scheme was used to perform the tomographic reconstruction [44]. The reconstructed volume is of size $512 \times 512 \times 512$ and contains about 9 virus particles, approximately 70 voxels in diameter. The approximate bounding box of the spike head (gp120) is $10 \times 10 \times 10$ voxels. The overall bounding box of the entire spike including the head of the spike (gp120) the tapering stem (gp41) and the adjoining virus envelope is $10 \times 10 \times 14$ voxels. The overall flow of the algo-

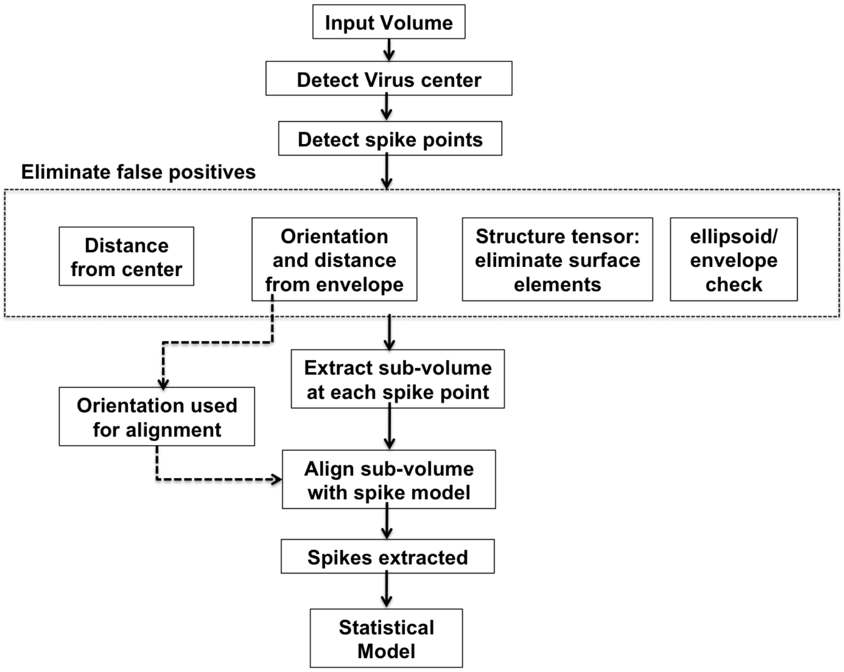


Figure 3.1: Algorithm flow of the spike detection and model generation method.

rithm is shown in Fig. 3.1. It begins by detecting the center of each putative virus particle then extracting a sub-volume that contains the virus particle. Candidate points that may lie on the spike are detected, and false positives are eliminated based on a number of specific physical criteria. Sub-volumes of the spike are extracted at each point, then are aligned and combined to create a statistical model.

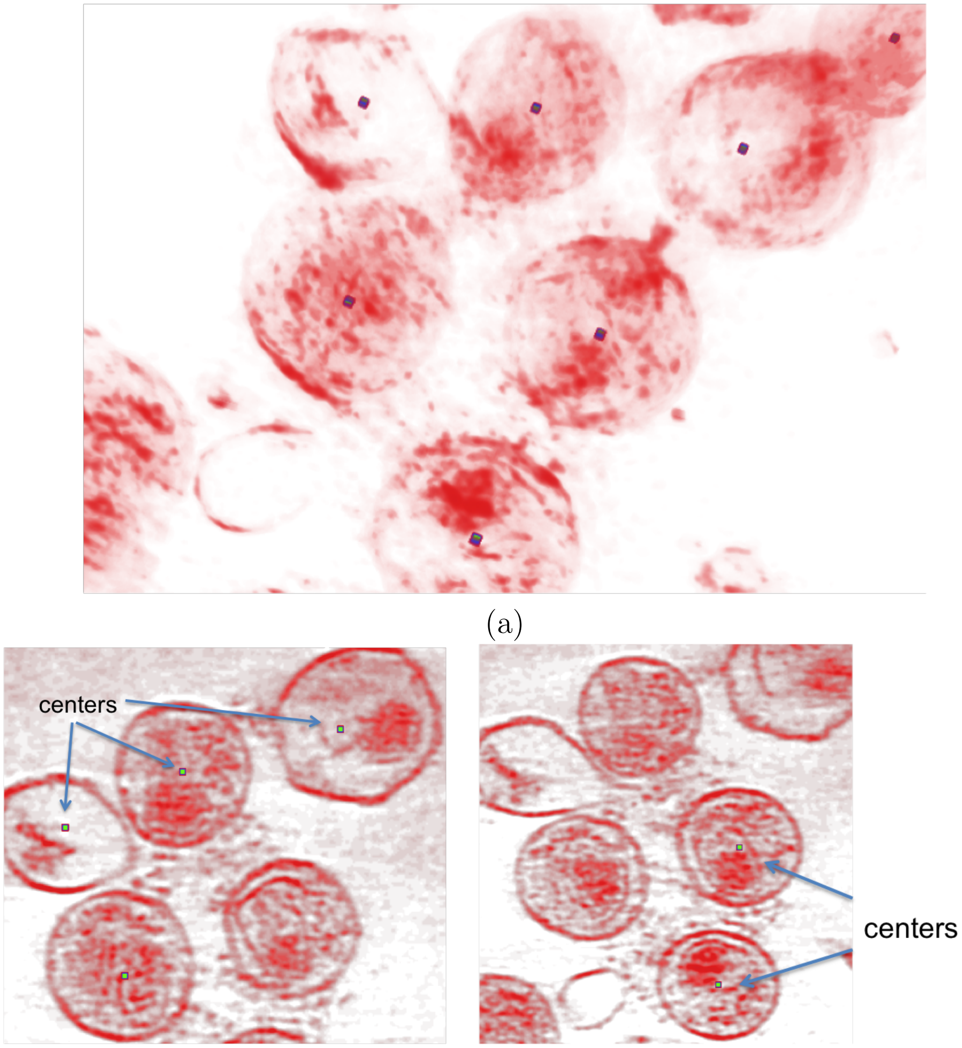
3.2.1 Spike Detection

3.2.1.1 Detect virus center

The first step of spike detection is to identify the centers of the virus particles. Since these have a roughly spherical outer envelope, a template matching technique was used to detect the spherical envelope region. 64 ellipsoid templates with radii varying from 33 to 36 voxels along the x, y and z dimensions were created. These ellipsoids were hollow with an outer shell of thickness 2 voxels corresponding to the width of the virus envelope. As a pre-processing step, the input volume containing the virus particles is thresholded with a very low value that eliminates low intensity background regions. The normalized cross correlation (NCC) is then calculated in the frequency domain for each of the ellipsoid templates with the input volume containing the virus particles. This results in 64 NCC volumes. The local maxima, in a $4 \times 4 \times 4$ region, over all of the 64 NCC volumes are selected as centers. Centers that lie near the volume edge are eliminated as false positives. The centers of all the virus particles were successfully captured using this method (Fig. 3.2). A sub-volume of size $100 \times 100 \times 100$ containing the entire virus particle centered on the detected virus center is then extracted from each data volume.

3.2.1.2 Detect spike-points

In this step candidate points that lie on a virus spike were detected in each virus sub-volume that was extracted previously. The head of the spike gp120 region is a blobby shaped structure. A difference of Gaussian



(a)

(b)

(c)

Figure 3.2: Input volume with detected virus centers (a) 3D volume rendering of the virus particles with the detected virus centers depicted as dots inside. (b) and (c) 2D slices: The centers of 4 virus particles occur on the same 2D slice and are shown in (b). The center of the fifth virus occurs on another 2D slice. The centers of 2 virus particles on the same 2D slice are shown in (c). The centers of the remaining virus particles are on other slice planes.

(DoG) operator was used to identify the blobby regions by selecting the local maxima of the DoG responses as candidate points. These are points that lie on blobby structures, including spikes. These candidate points are referred to as *spike-points*. The DoG is a close approximate of the second derivative of a Gaussian (Laplacian of Gaussian). Evaluating the DoG involves subtracting two different scales of the sub-volume region of the virus particle. A scale space of 5 volumes was created by convolving the original with a Gaussian kernel at $\sigma = [0.707, 1.41, 2.12, 2.828, 3.355]$. Four DoG volumes are generated by subtracting two consecutive scales, i.e. volumes at $\sigma = 0.707$, 1.41 are used to generate one DoG volume and $\sigma = 1.41$, 2.12 are used to generate another and so on (Fig. 3.3).

Local Maxima of DoG: Local maxima are located for each DoG volume and its immediate neighbors. The local maxima check is performed for each voxel's 26 neighbors in the current DoG volume and the DoG volume above and below it (see Fig. 3.3). A voxel in DoG-2 is compared against its 26 neighbors and the 27 voxels in the corresponding location of DoG-1 and DoG-3. Similarly DoG-3 voxels are compared with DoG-2 and DoG-4. The resulting local maxima points are referred to as spike-points and are candidate points on the spikes. Typically, about $1,000$ spike-points were obtanined for a single virus particle. The next steps attempt to eliminate the false positives and preserve only those points that lie on a spike.

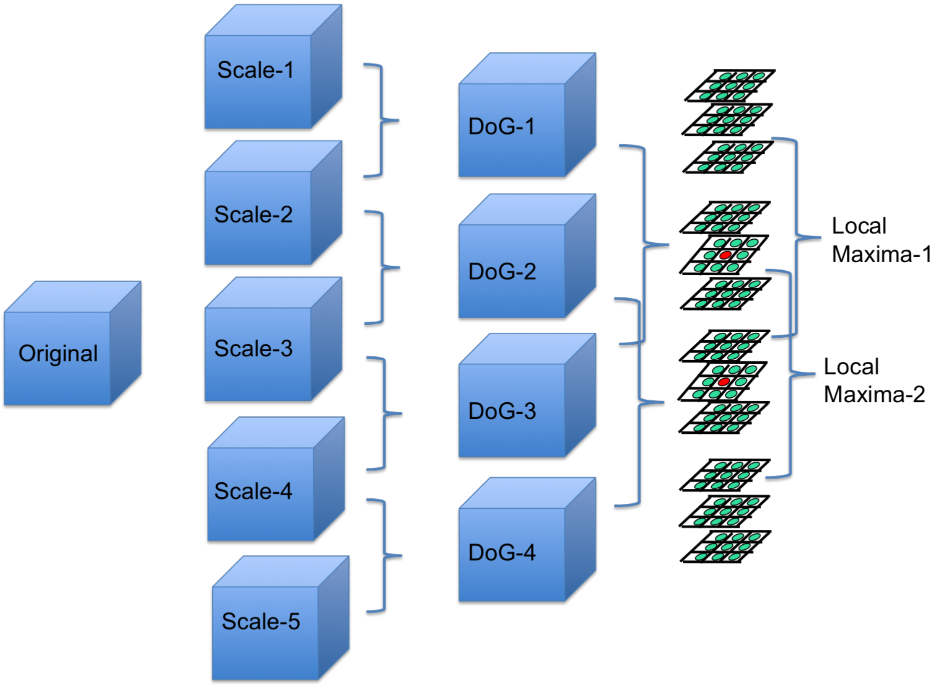
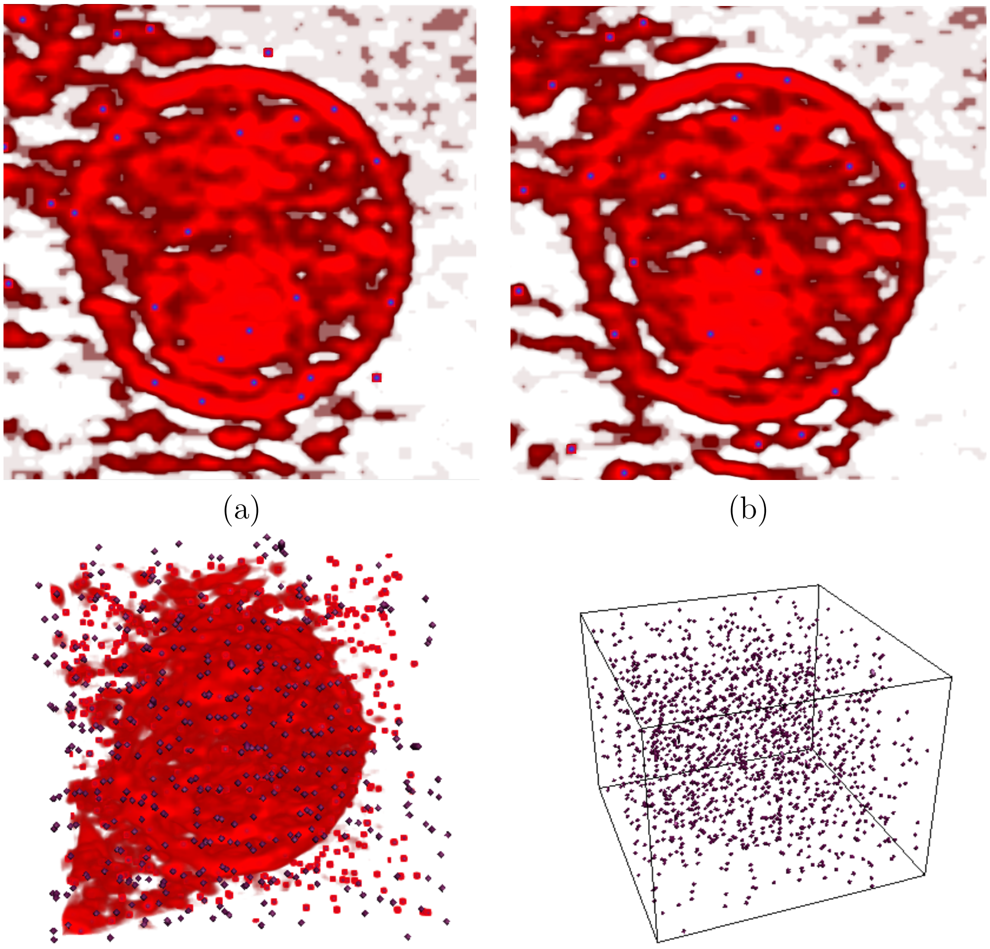


Figure 3.3: Detecting spike-points: The subvolume containing the virus particle is scaled by convolution with a Gaussian kernel at different sigmas. Difference of Gaussian (DoG) volumes are computed by subtracting neighboring scaled volumes. Local maxima of each DoG volume including the local neighborhood of the adjoining DoG volumes are identified. These local maxima are points which lie on blobby regions of the original volume and are called spike-points.

3.2.1.3 False positive reduction

An array of stages was used to eliminate false positives from the detected spike-points. A soft threshold approach was used, by defining a confidence range $[0, 1]$, where 1 indicates high confidence. Confidences are assigned to each spike point at every false positive reduction stage. The decision on whether a point is a false positive is made at the end by combining the confidence values from all the stages.



(a) (b)

(c) (d)

Figure 3.4: Spike-points before FP removal: (a) and (b) are 2D slices of a virus particle with spike-points shown in blue. (c) is a 3D volume rendering of the virus particle with spike-points. (d) shows the spike-points in 3D in a sub-volume containing a virus-particle (not shown).

Distance from virus center Based on the current literature about the structural characteristics of the AIDS virus and the typical virus radii seen in the data, it was observed that spikes on the virus envelope occur at least 30 voxels from the approximate virus center. Spike-points that lie less than 30 voxels from the virus center are most likely false positives. A soft-threshold value of 1 was assigned for all spike-points further than 30 voxels from the center. Those that are below 30 are assigned a confidence value of $1-\frac{30-distance}{30}$. Points that are very close to the center, at about 20 voxels, are rejected.

Distance from envelope and orientation Given a spike-point, the orientation of the spike-point was calculated with respect to the virus envelope and estimated its approximate distance from the virus envelope. Spikes typically protrude radially from the virus envelope (Fig. 3.6). As explained in the description of the virus center detection (Section 3.2.1.1) the algorithm finds the ellipsoid that best correlates with each virus particle. Given the ellipsoid parameters, the normal from the surface of the ellipsoid to the spike point is calculated. Let $\vec{p}$ be the spike-point and S the ellipsoid surface. Then the vector $\vec{x}$, such that the spike axis $\vec{px}$ is normal to the ellipsoid surface S at $\vec{x}$ (Fig. 3.5) satisfies

$$(\vec{p}-\vec{x})\vec{x}'=0, \tag{3.1}$$

where $\vec{x}'$ is the tangent at $\vec{x}$ and $\vec{p}-\vec{x}$ is the normal. The vector $\vec{x}$ may be parametrized in spherical coordinates for computational efficiency [31]:

$$x(\vec{\theta},\phi) = r\begin{bmatrix} \alpha cos(\phi)sin(\theta) \\ \beta cos(\phi)sin(\theta) \\ \gamma sin(\phi) \end{bmatrix} \tag{3.2}$$

where α, β, γ and r represent the parameters of the ellipsoid.

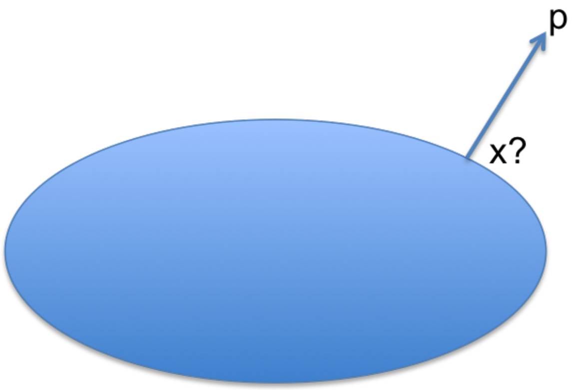


Figure 3.5: Spike orientation axis: Given a spike-point $\vec{p}$, the objective is to find $\vec{x}$ such that the spike orientation axis $\vec{px}$ is normal to the surface of the ellipsoid.

Proceeding directly, solve for θ and ϕ in the set of equations

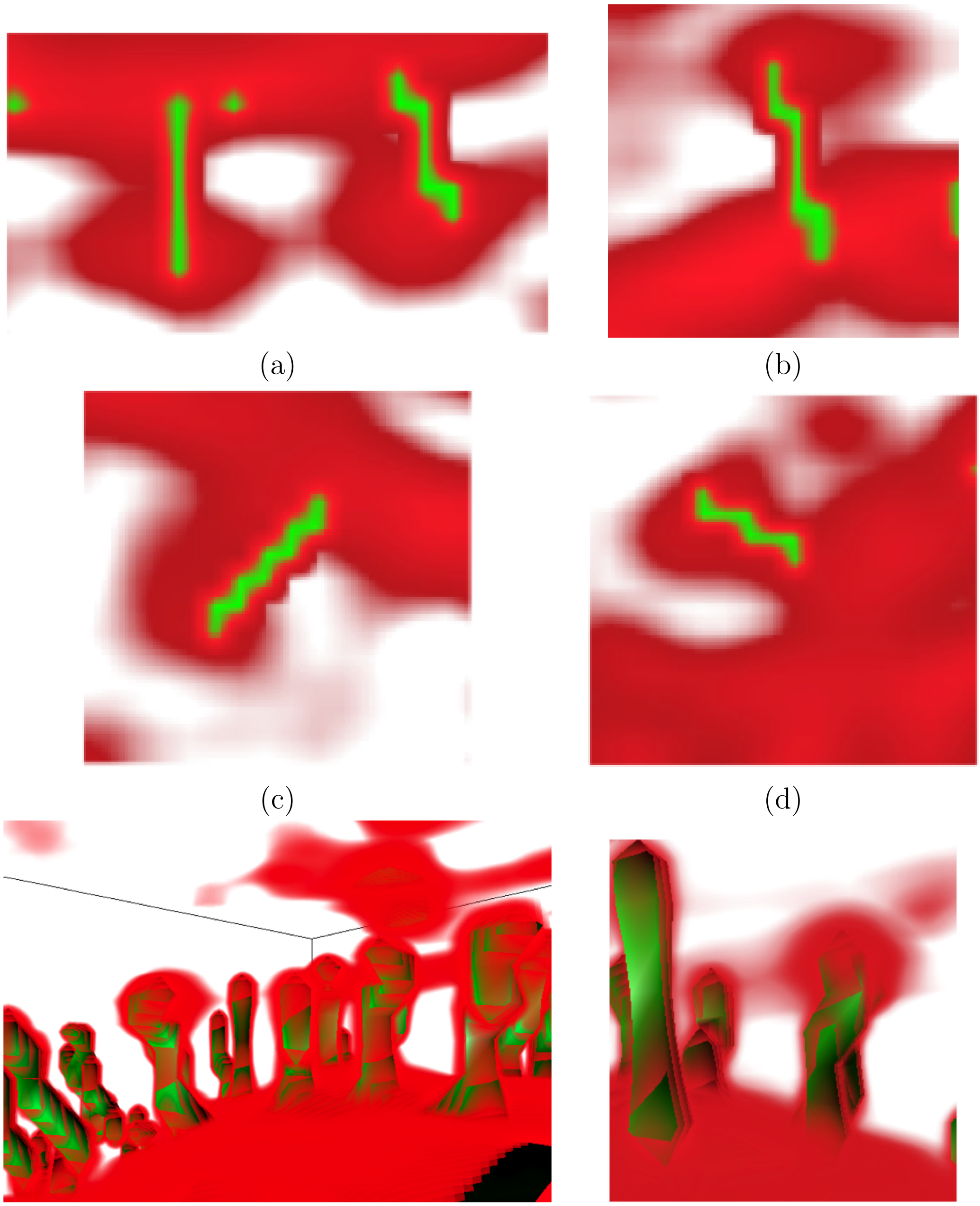
$$f := \begin{cases} (\vec{p} - \vec{x})\frac{\partial \vec{x}}{\partial \theta} = 0 \\ (\vec{p} - \vec{x})\frac{\partial \vec{x}}{\partial \phi} = 0, \end{cases} \tag{3.3}$$

by using Newton's method:

$$\begin{bmatrix} \Delta\theta \\ \Delta\phi \end{bmatrix} = -J^{-1} f, \tag{3.4}$$

where J is the Jacobian of f.

The iterations begin with an initial estimate $\theta = tan^{-1}(\frac{\alpha p_x}{\beta p_y})$ and $\phi = tan^{-1}(\frac{p_z}{c\sqrt{(\frac{x}{a})^2+(\frac{y}{b})^2}})$ and are updated with $\Delta\theta$ and $\Delta\phi$. Convergence is obtained in about 3 iterations. The θ and ϕ parameters found using Newton's method was used to estimate x as given in (3.2). Using this formulation, a likely orientation axis of each spike $\vec{px}$ and its length $|\vec{px}|$ is computed. Note that



(a) (b)

(c) (d)

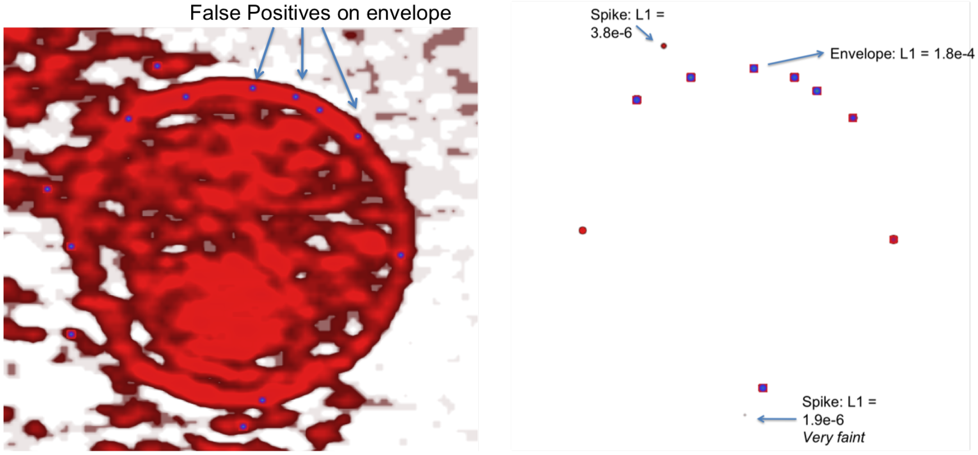
(e) (f)

Figure 3.6: Spike orientation axis: (a, b, c, d) are 2D slices with the detected spike orientation axis shown in green. (e, f) are 3D volume renderings of the spike region with the detected axis rendered in green.

the actual orientation of the spikes may differ slightly from the calculated orientations since the virus envelope is approximated by an ellipsoid and the spike orientation may not be exactly normal to the envelope surface. The orientation axis estimated using this approach is a good initial estimate for the alignment and spike extraction performed in Section 3.2.1.4. The length $|\vec{px}|$ is used to eliminate spike points which lie too far away from the surface. A confidence value of 1 is given to spike-points that are less than 10 voxels in length. If the distance is greater than 10 then a confidence value of $1 - \frac{|length-10|}{10}$ is assigned. Spike points that are greater than the length threshold by a tolerance (here 25%) are eliminated. A fast check to eliminate spike-points that lie inside the ellipsoid and hence the virus is performed. The distance of the spike point $\vec{p}$ from the virus center is compared with the distance of the ellipsoid point $\vec{x}$ from the center. If the distance of $\vec{p}$ to the center is smaller than $\vec{x}$ to the center, then it implies that the point lies inside the ellipsoid and is eliminated.

Fitting ellipsoid on envelope The ellipsoid template was used to detect false positive spike points that lie on the virus envelope. For each virus particle, the ellipsoid that gave the highest NCC value for its center (Section 3.2.1.1) is chosen as a template for the virus. Affine registration is performed between the ellipsoid and the corresponding virus to align them. Spike points that lie on the registered ellipsoid or inside it are tagged as false positives. A false-positive membership value is then assigned based on the fraction of neighbors of the spike-point that are inside the ellipsoid.

Structure tensor Spike-points that lie on the virus envelope were a large source of false positives. Structure tensors are used to detect points that lie on the envelope. At each spike-point, the structure tensor is calculated and points that were on a surface-like structure are eliminated while preserving those that lie on blobby regions. To make the structure tensor calculation more robust, local region growing is performed in a $5 \times 5 \times 5$ region around the spike-point, thereby enabling the computation of partial derivatives on the points extracted.



(a) (b)

Figure 3.7: Structure tensor for false positive elimination: (a) Shows false positives on the virus envelope that are eliminated by evaluating the structure tensor. (b) Points on the envelope have a large L_1 value compared to the ones on the spike, which are blobby. False positives are eliminated based on this.

$$
\begin{aligned}
[I_x, I_y, I_z] &= \nabla(\vec{p}) && (3.5)\\
StructureTensor &= \begin{bmatrix} I_x^2 & I_xI_y & I_xI_z \\ I_xI_y & I_y^2 & I_yI_z \\ I_xI_z & I_yI_z & I_z^2 \end{bmatrix} && (3.6)
\end{aligned}
$$

$$[L_1, L_2, L_3] = Eigen(StructureTensor) \quad (3.7)$$

Here I_x, I_y, I_z are the partial derivatives at spike point $\vec{p}$ and L_1, L_2, L_3 are the eigen values of the structure tensor. Spike points that lie on the envelope of the virus have a surface-like neighborhood with a "surfel" (Surface-Element) characteristic, where:

$$L_1 \gg L_2 \approx L_3 \quad (3.8)$$

L_1 for points that were on the envelope $\geq 10^{-4}$ while those that were on spikes are observed to be $\leq 10^{-6}$. L_2, L_3 are much lower, on the order of 10^{-8}. Points with $L_1 \leq 2 \times 10^{-4}$ (threshold) are assigned a confidence level of 1 while those that are greater than the threshold are assigned a value of $1 - \frac{|L_1 - threshold|}{threshold}$. Points with a value beyond the threshold by a tolerance of 25% are eliminated as false positives. Points on spikes, which are blobby and have much lower L_1, are preserved.

At this stage the algorithm has eliminated a large percentage of false positives, yielding about 150 spike points and a set of membership values for each point based on the false positive elimination stages. The membership values from each stage are averaged and assigned as the final membership value for each spike point. The final membership values of most spikes are close to 1 with about 20% of the points having membership values ranging from 0.75 *to* 1. A Free-response Receiver Operating Characteristic (FROC) was plotted to estimate the optimum choice of threshold of the final membership value to eliminate false positives (Section 3.3.1).

3.2.1.4 Extracting spikes

The next step was to extract spikes at the detected spike points and align them to build a statistical model.

Phantom spike A phantom structure shaped like a virus spike was created to aid in extracting the virus spikes. A cylindrical structure with a sphere placed on top is created to replicate the shape of a virus spike. This model is blurred with a Gaussian. The blur parameters are estimated by simulating projections of a sphere between $\pm 69^\circ$ in steps of 3° and reconstructing it through back projection. The resulting sphere with blur due to limited angle tomographic effects is used to estimate the σ value used for blurring the spike model (Fig. 3.8). For each spike point, the extracted phantom spike is aligned along the spike axis predicted in Section 3.2.1.3.

Extract subvolume and its orientation At each spike point that has filtered through the false positive removal process, a $10 \times 10 \times 14$ sub-volume, the observed size of a typical spike (see Section 3.2), is extracted. The sub-volume's intensity range is normalized to lie in the interval $[0, 1]$ and processed using thresholding and connected component analysis. The choice of threshold varies from $0.45 - 0.2$ where the appropriate threshold is selected based on the FROC analysis (see Section 3.3.1). At high threshold values, spike points that lie on blobby regions with very weak intensity regions can break up into several small connected components. These spike-points are eliminated as false positives as a reliable spike region cannot be extracted. While spike points that lie on spikes with good contrast and that are distinct from the background produce a large connected component that includes the spike-point, these are preserved. This sub-volume is next compared with the phantom spike model in order to recover its orientation.

An affine registration between the sub-volume region and the phantom spike oriented along the predicted spike axis is performed. This is to recover only small rotation and translation shifts between the predicted axis and the actual axis of the spike. Large rotations that would invert the orientation of the spike are thereby prevented. With this, the orientation of the spike present in the sub-volume was estimated.

The spike point is shifted by ± 1 along each dimension and the spike sub-volume extraction process described above is repeated on each shifted spike-point. The sub-volume that delivers the best similarity match with the phantom model is selected. A wavelet based structure similarity index developed by Sampat et al. [41] was used to compute the similarity between the sub-volume and the phantom model. Structure similarity measures provide more robust matching than metrics such as mean square error.

This procedure described above was performed for all the spike points and the sub-volume containing the spike was extracted and its orientation was determined(Fig. 3.9). With the orientation established, it is now possible to combine all the extracted spikes in the next step of building a statistical shape model.

3.2.2 Building a Statistical Shape Model

The spikes extracted from the previous step (Section 3.2.1.4) are all aligned in a common coordinate frame. The spikes in the AIDS virus exhibit three-fold symmetry about the central axis. The pose of each spike is recovered by rotationally aligning it with the spike model used by Liu et al. [25]. This is done only to recover the pose or rotation of the spike along the axis of the spike. At this stage all the spikes are fully aligned with each other.

To build the statistical model, the information from all the aligned individual spikes is combined to form a probability density function at each voxel. The statistical model is in 4D space where the fourth dimension is the probability density function. The density function was constructed at each voxel based on the intensity values of the detected spikes at that voxel. A kernel smoothing density estimate was used to construct the density function at each voxel.

3.2.3 Spike Membership and Model Refinement

Each detected spike's voxel has a membership value in the overall statistical model. A membership volume for each spike was built, where each corresponding voxel is assigned an intensity value equal to the membership of that voxel in the statistical model (Fig. 3.9). Noisy regions of the spike have low membership values. An average membership number for each membership volume was reported. This gives an overall membership value of a spike in the statistical model. It was noticed that spikes that are blurry and noisy have lower average membership values while those that are more distinct have higher average membership values.

3.3 Results and discussion

3.3.1 FROC Analysis

To analyze the spike detection algorithm, a Free-response Receiver Operator Characteristic (FROC) study [5] was performed. A FROC curve is a plot of sensitivity vs the number of false positive detections per virus. Sensitivity is defined as the fraction of spikes detected.

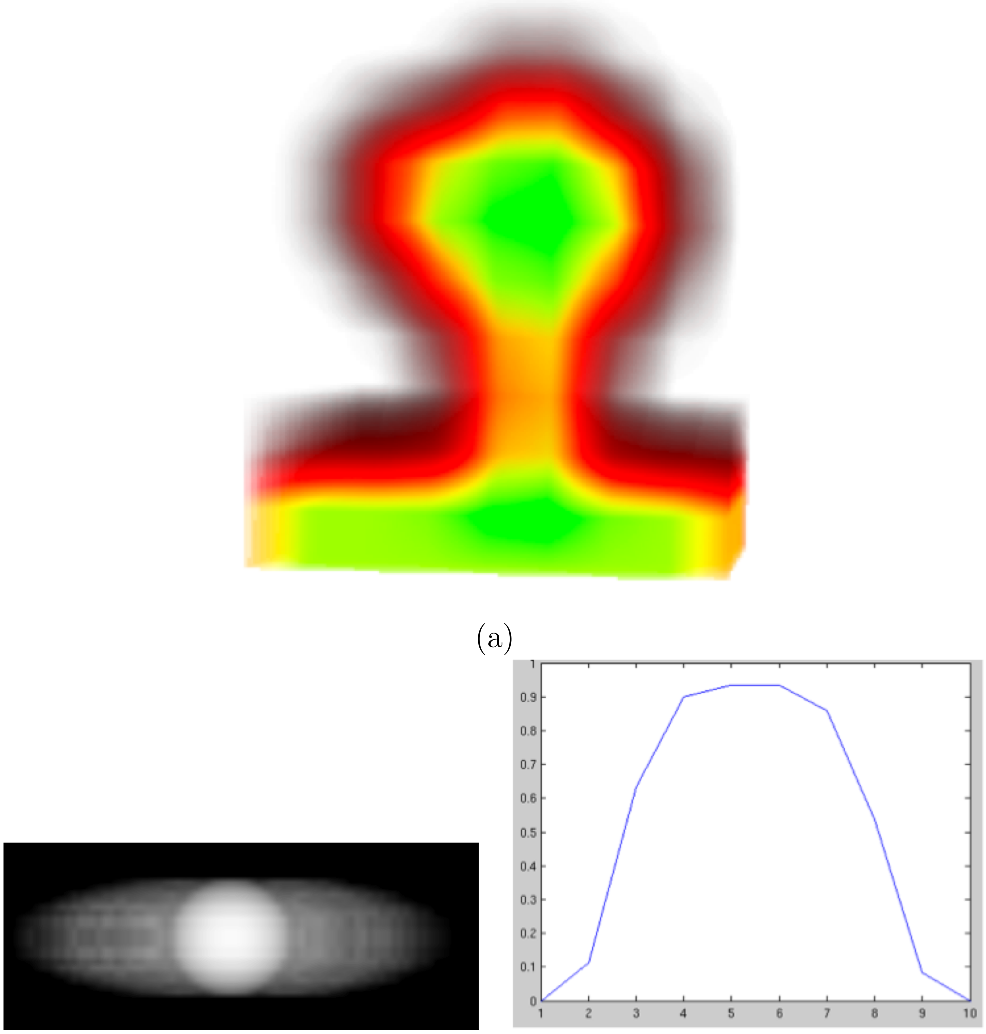
$$Sensitivity = \frac{Number\ of\ True\ Positives}{Total\ number\ of\ spikes} \tag{3.9}$$

For this study, ground truth for 96 spikes in four different virus particles were marked. A knowledgeable user marked a point on the head of the spike's approximate center. The detection and extraction algorithm was run and the resulting location of spike-points compared against the ground truth. A spike was considered detected if a spike-point was placed by the algorithm within a radius of 7 voxels from the marked ground truth point (acceptance radius). This was chosen because the head of the spike gp120 region is about $10\times10\times10$ voxels (see spike description in Section 3.2), which corresponds to a midradius (center to edge) of about 7 voxels. Multiple spike-point detection within this radius was considered as a single detection. The detection algorithm's final membership value and the threshold value used in the extraction of spikes were varied and the resulting sensitivities was plotted against the false positive rate.

The best sensitivity was at 0.81 where 77 out of 96 spikes were detected with 9 false positives (FP) per virus. At 0.75 sensitivity, 72 out of 96 spikes were detected and only 7 FP per virus. This seems to be the best operating range for the detection algorithm. Below this point, the sensitivity drops sharply. When the false positive elimination stages was completely relaxed, 99% of the spikes were detected. But this also yielded a large 56 false positives per virus. At 93% sensitivity 26 false positives were found per virus. The best operating range is at a sensitivity of about 75 to 80% with 7 to 9 false positives per virus.

3.3.2 Statistical Shape Model

The statistical model is in 4D data where the fourth dimension represents the probability density function associated with each voxel. Visualizing and displaying such data in its entirety is challenging. Fig. 3.11 shows a 3D volume rendering and a central 2D image along the XZ plane of the mean sta-



(a)

(b)

(c)

Figure 3.8: Phantom model used for extracting a spike: (a) Blurred spike model built using ellipsoid on top of a cylinder and an envelope region at the base. (b) 2D slice of reconstructed (backprojection) sphere phantom, used to estimate blurr parameters. A Gaussian blur with the estimated parameters is applied to the ellipsoid-cylinder model. (c) 1D intensity profile through the center of the 2D slice shown in (b).

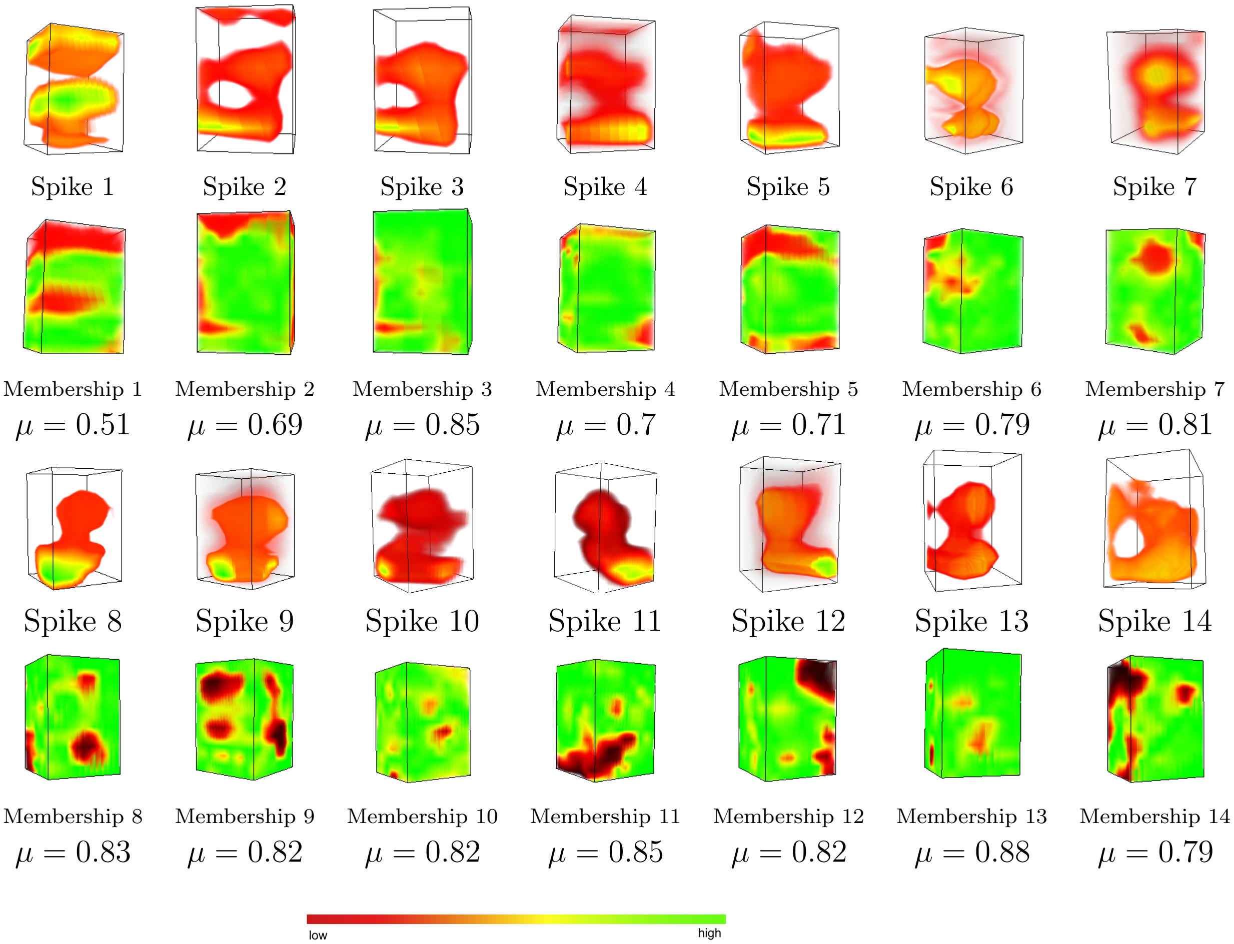


Figure 3.9: Automatically extracted spikes and their membership volumes according to the statistical model. The statistical model has a probability distribution associated with each voxel. A membership volume where each voxel in the membership volume corresponds to the membership value of the spike's voxel in the statistical model was computed. The mean membership value for each membership volume is also shown. Noisy regions of the spikes tend to have lower membership values at those regions.

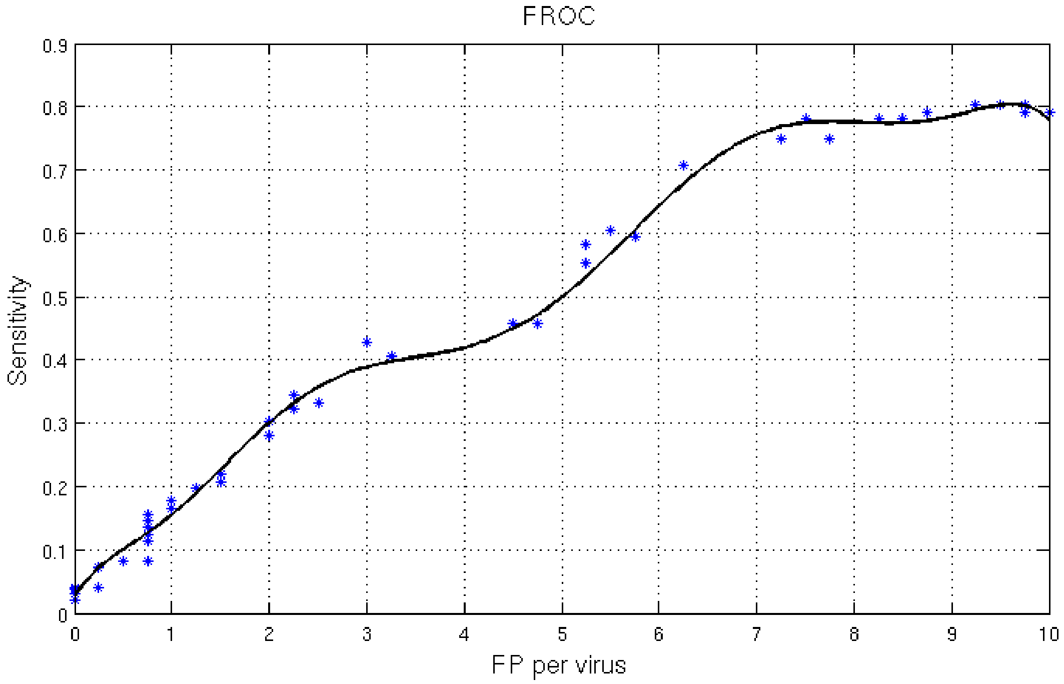
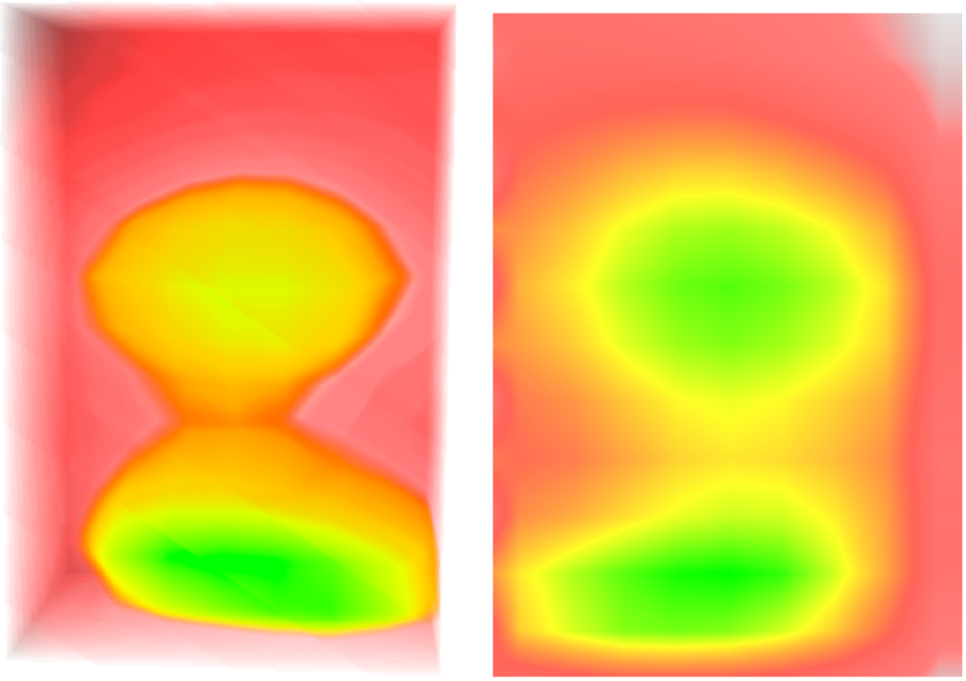
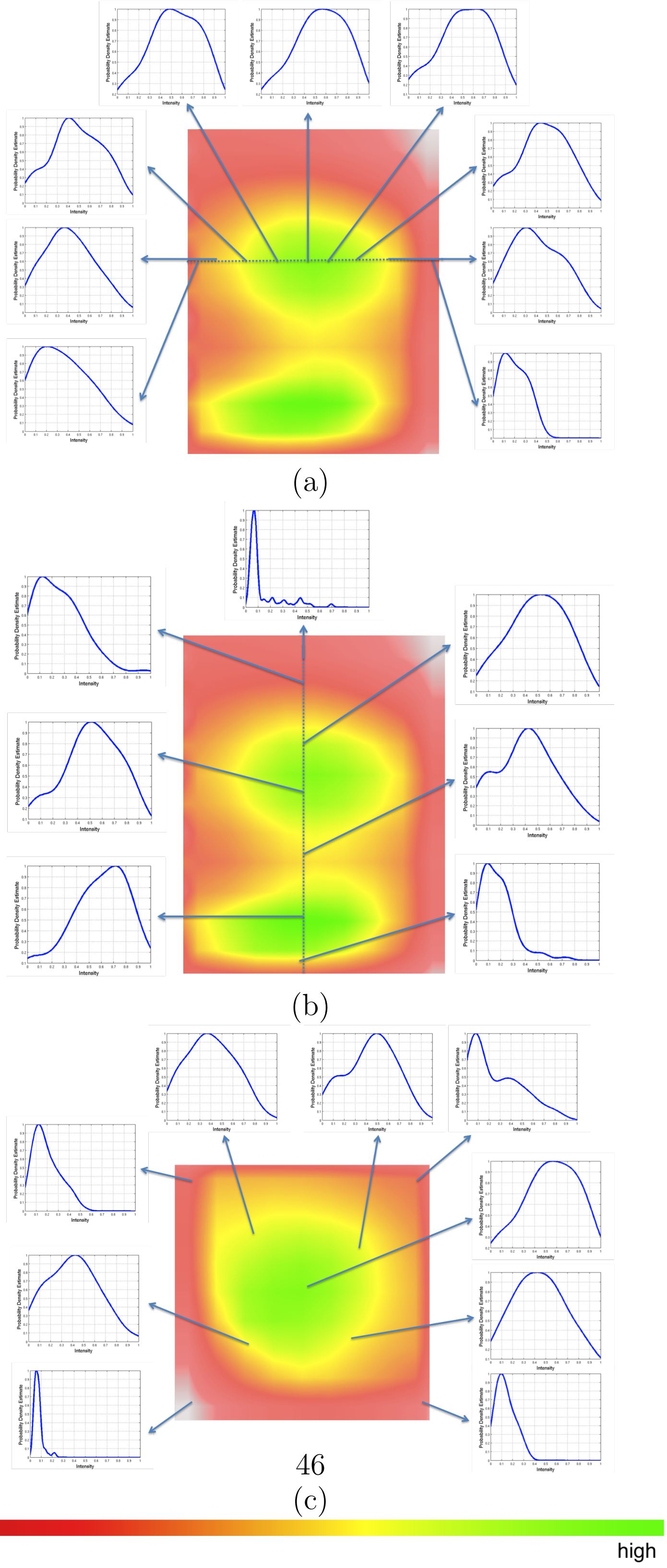


Figure 3.10: Free response Receiver Operating Characteristic (FROC) curve plots the sensitivity vs the number of false positive detections per virus. 80% sensitivity (80% of all spikes were detected) was measured with about 9 false positives (FP) per virus. The best operating range is at 7 FP with a sensitivity around 75%, beyond which the sensitivity drops.



(a) (b)

Figure 3.11: (a) Volume rendering (VR) of the mean statistical shape model of the AIDS virus spike. (b) is a 2D slice of the mean statistical shape model.



(a)

(b)

(c)

Figure 3.12: Statistical Shape Model: (a) and (b) show a profile image of the mean statistical model along the XZ plane and plots of the density function associated with voxels that lie on the center line shown in the image. (c) XY plane. Voxels that are at the center have higher mean values (green), while those near the edges have lower mean values (red).

tistical model. Each voxel in the mean statistical model has an intensity value that is the mean value of the detected spikes at that voxel. 2D profiles of the average spike model are displayed and the density function on a set of voxels was plotted. Figs. 3.12(a) and (b) show a profile image of the average statistical model along the XZ plane and plots of the density function associated with voxels that lie on the center line shown in the image. Fig. 3.12(c) is an image along the XY plane of the average statistical model and the associated density function plots are shown. Voxels that lie near the center of the image have a bigger spread in their density function and larger mean values than those that lie near the boundary. A small spread for voxels further from the center along with a low mean value implies a high confidence (low uncertainty) bound for the size and shape of the spike. As seen in Fig. 3.12(b), there is a higher uncertainty associated with the stem region (gp41) that connects the head of the spike (gp120) to the envelope. Whereas, the head of the spike region (gp120) and the envelope have greater confidences and higher mean values.

3.3.3 Applications

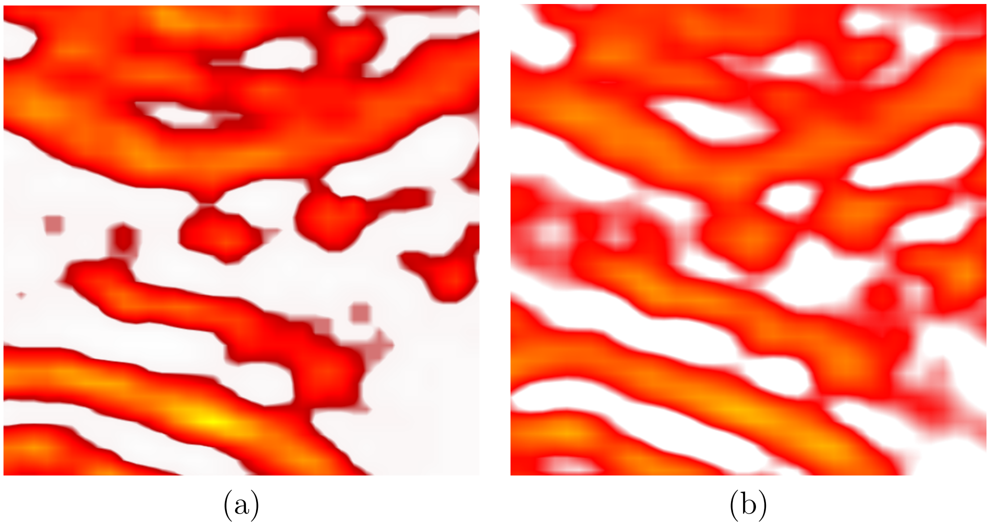
The statistical model can be used for various computational biology applications. Some important applications are discussed here:

3.3.3.1 Fitting

Using X-ray crystallography studies, biologists can determine the atomic structures of the subcomponents of biological complexes. A drawback is that the biological complex is not in its native state since these atomic structure models are obtained using crystal lattice growing techniques. A common practice is to perform fitting [37], where the atomic model is aligned with a coarser model obtained from EM imaging, in which the biological complex is in its native state. Fitting is a difficult problem where it is necessary to explore all possible angles and orientations of the atomic model that can fit the EM model. Typically, a single EM model is used for fitting. A statistical EM model can provide a better range of fitting results where the atomic model can be fitted into one of several possible spike shapes, providing more flexibility. Such powerful modeling can further aid in providing more accurate atomic models of critical subcomponents of biological complexes such as the spike region of the AIDS virus.

3.3.3.2 Computer aided drug design

Development of computer algorithms for bio-simulations have enabled extensive studies of the protein structure and dynamics of new potential drug targets [45]. Numerous docking programs are extensively used in the biotechnology and pharmaceutical industries. The algorithms for docking use force-field-based methods such as molecular dynamics or Monte Carlo simulations which allow for movements of ligands and targets [45]. Most docking programs assume the structure to be rigid with a single shape. Using statistical shape models instead of a single shape model, docking and other computer aided drug design results could be further enhanced. Shape complementarity based methods fit the ligand shape into the negative shape of the protein structure. Using the statistical shape model of the virus spike, shape complementarity tests can be performed on candidate drugs using a wide range of possible spike shapes with associated statistical parameters.



(a) (b)

Figure 3.13: Tomographic Reconstruction using (a) Shape based regularization, where the mean spike model was used as the prior shape in the reconstruction process, (b) Weighted Back Projection reconstruction. The shape based reconstruction method shows reduced blurring and improved spike feature visualization.

3.3.3.3 Shape based tomographic reconstruction

A new tomographic reconstruction introduced in chapter 4 uses known shape models of certain critical structures in the tomographic reconstruction process. The shape models are incorporated into the tomographic reconstruction through a regularization process and a MAP (maximum *a posteriori*) estimate is obtained. Local segmentation of features is performed at each iteration of the reconstruction process and compared with the prior shape model. Over or under-segmentation detected at each voxel of the local feature drives a scaling factor of the regularization term.

The mean shape of the statistical spike model was used as the prior shape in the regularization process of the reconstruction method. The resulting reconstruction shows reduced blur and improved feature visualization (Fig. 3.13). Further enhancements could an include a full Bayesian reconstruction scheme that completely utilizes the 4D statistical spike model in the reconstruction process as the prior probability distribution. Improved Electron Tomography reconstruction will provide better structure visualization giving important biological information around the vicinity of critical structures like the virus spikes.

Chapter 4

Shape Based Regularization of Electron Tomography Reconstruction

A statistical approach for solving inverse problems using a Bayesian framework has been used before for a variety of reconstruction problems. Typically all variables in statistical inverse problems are treated as random variables. The estimate of the posterior probability distribution is the solution to the inverse problem. Under a Bayesian framework, the prior information is encoded by a probability distribution and is called the *a priori* (or simply, prior) [21].

Bayes rule gives us the relation

$$p(f|g) = \frac{p(g|f)p(f)}{p(g)}, \tag{4.1}$$

where $p(f)$ is the prior about f, $p(g|f)$ is the likelihood term that models the physics of image acquisition. $p(g)$ is the marginal probability distribution that is typically ignored when using estimators since it does not vary with f [10] [21]. Also, calculating $p(g)$ is computationally expensive and is required only if a complete distribution of the posteriori needs to be constructed. Estimators make the reconstruction more computationally feasible as the marginal $p(g)$

can be ignored [21].

Various priors have been used by different groups depending on the imaging model and application. In medical imaging, Gindi et al. [10] have used edge maps of anatomical images as priors while performing reconstruction of functional images. Their goal was to induce edges and hence better contrast in functional images (PET, SPECT). The anatomical image (typically CT, MRI) is registered with the functional image (typically PET, SPECT). An edge map is created using the anatomical image which is used as a prior while performing reconstruction of the functional image. Gindi et al. model both the likelihood and prior as Gibbs distribution. A predetermined edge map is built from a CT or MRI image and is incorporated in the a priori as an energy term. Anatomical information from MRI and CT have similarly been used in Hsiao et al. [20] and Rangarajan et al. [35].

Methods that use global smoothness priors do not incorporate structure information in the reconstruction process. The assumption of smoothness can be arbitrary and can affect edges. Methods like Gindi et al. [10] and Hsiao et al. [20] that use CT/MRI image as priors for PET reconstruction have issues where edges seen in a CT/MRI image are not always seen in a PET image leading to inconsistent features.

4.1 Statistical Regularization

The statistical regularization framework allows us to use shape based priors explicitly in the reconstruction process. Also, the physical process of data collection can be modeled through the likelihood term.

4.1.1 Likelihood probability distribution:

A Poisson distribution is used in a system that has a high number of events of which only a small percentage are counted. In the context of ET, a large number of electrons are bombarded on the specimen and a large fraction is scattered from the small collection aperture of the EM creating contrast. As already mentioned, the dominating contribution to the noise is shot noise which is Poisson distributed. Other modalities like SPECT and PET also use a Poisson distribution to model noise arising during the image acquisition [13], [52], [10]. Under this assumption, the likelihood becomes

$$p(g|f) = \prod_{m}^{M} \frac{(Hf)_m^{g_m} \exp(-(Hf)_m)}{g_m!}, \tag{4.2}$$

where M is the total number of projection measurements, i.e. size of a 2D projection image times the number of projections, g_m is the m^{th} projection measurement and $(Hf)_m$ is the corresponding projection of the current reconstruction estimate f using the forward operator H.

4.1.2 Prior probability distribution

The prior $p(f)$ is modeled as a Gibbs distribution:

$$p(f) := \exp\left(\frac{-\lambda S(f)}{Z}\right), \tag{4.3}$$

where $S(f)$ is an energy functional, Z is a normalization factor and λ is a weight. Gibbs distributions are widely used in image restoration problems and have a number of advantages. They can enforce smoothness in images as well as model images containing piecewise homogeneous regions [23]. They also satisfy the Markov property, i.e. only neighboring voxels have direct interactions with each other (Markov-Gibbs equivalence).

4.1.3 MAP estimate and optimization

Maximum-a-posteriori (MAP) is one the most popular estimators used in statistical regularization [21] [10]. The MAP estimator is defined as the element f that maximizes the posteriori $p(f|g)$, alternatively minimizes the negative log of the *a posteriori*. An optimization step is involved in calculating the MAP estimate. Gindi et al. [10] use a EM-ML optimization step to solve the objective function resulting from their MAP estimate. Hsiao et al. [20] use deterministic annealing for optimizing the MAP. Their resulting objective function is not convex since they use a Poisson likelihood probability distribution and a Gamma prior probability distribution.

A Poisson likelihood assumption is combined with a Gibbs prior to perform a segmentation-based optimization with an EM algorithm. The MAP estimator is given as the solution to:

$$f_{\mathrm{MAP}} := \arg\max_{f} p(f|g), \tag{4.4}$$

A cost function $c(f)$ is defined as:

$$c(f) := \log(p(g|f)) + \log(p(f)), \tag{4.5}$$

where $\log(p(g|f))$ is the log likelihood and $\log(p(f))$ acts as a regularizer. $\log(p(f))$ is minimum when the energy term $S(f)$ is minimum. Using Bayes theorem with the prior in (4.3), the MAP estimator given in (4.4) can be written as (1.4).

It can be shown that under certain circumstances, (1.4) can be solved using iterates from a penalized EM method [30] [14].

$$f^{(i+1)} := \frac{f^{(i)}}{1 + \lambda \left| \nabla S(f^{(i)}) \right|} H^T \frac{g}{Hf^{(i)}}, \tag{4.6}$$

where $S(f^{(i)})$ acts as the regularizer, λ is the regularization parameter, and H^T is the adjoint of the linear forward operator, which in tomography is the back-projection operator. Note that the division above between functions is taken point wise, and if functions are represented by vectors (discretized), it is taken component wise.

4.2 Shape Based Regularization

A regularization approach is introduced in the framework of statistical regularization, where the energy term $S(f)$ encodes local information about the structure being reconstructed f. Spatial models M of local structures (virus spikes in the example that follows) are built using PDB and EMDB information (see Section 4.3). Models can also be built using a composite segmentation technique [11] where virus features segmented from different reconstructions are aligned together and combined through a voting scheme to provide a high confidence model. In this reconstruction method, at each iteration virus spikes of the current reconstruction are segmented creating a binary volume (0 or 1) that is compared with the known spatial model of the structure. The regularizer applied at a voxel in f_j is chosen to be either a penalizing or a increasing term depending on whether the voxel neighborhood is determined to be over-segmented or under-segmented, respectively.

4.2.1 Local comparison of sets and under- & over-segmentation

For a given measurable domain $U \subset \mathbb{R}^3$, let $\chi_U \colon \mathbb{R}^3 \to \{0,1\}$ denote an indicator function of U, i.e

$$\chi_U(x) := 1 \text{ if } x \in U \text{ and } 0 \text{ otherwise.}$$

Let $|U|$ denote the Lebesgue measure of U, so

$$|U| := \int_{\mathbb{R}^3} \chi_U(x)\, dx.$$

Finally, let $B_r(x)$ denote the ball in $\mathbb{R}^3$ with center at x and radii $r > 0$.

Consider two domains $M, U \subset \mathbb{R}^3$ and let $r > 0$ be fixed. For $x \in \mathbb{R}^3$, define

$$\triangle_r(U,M)(x) := \frac{\big|\{y \in B_r(x) : \chi_M(y) = \chi_U(y)\}\big|}{\big|B_r(x)\big|} \tag{4.7}$$

Hence,

$$0 \le \triangle_r(U,M)(x) \le 1 \quad \text{for all } x \in \mathbb{R}^3 \text{ and all } r \ge 0$$

which measures the relative portion of a local r-neighborhood of x where the sets U and M agree. To introduce the concept of under- & over-segmentation, consider $x \in \mathbb{R}^3$ with $r > 0$ fixed and $\epsilon > 0$. One of the sets, M, is considered as the given *shape model*. Then, U is *under-segmented* at x compared to M if

$$\chi_M(x) > \chi_U(x) \text{ and } \left|U \bigcap B_r(x)\right| > \epsilon.$$

Correspondingly, U is *over-segmented* at x compared to M if

$$\chi_M(x) < \chi_U(x) \text{ and } \left|U \bigcap B_r(x)\right| > \epsilon.$$

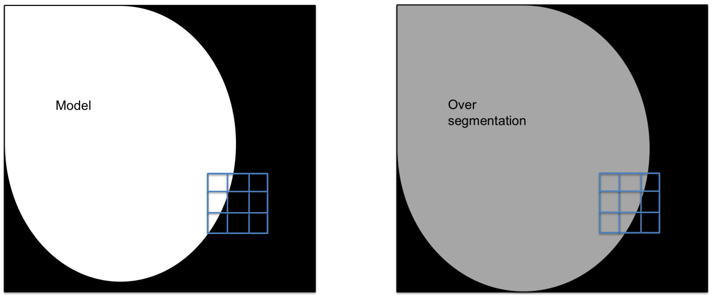
The parameter $\epsilon > 0$ is a tolerance level. Define

$$s_{r,\epsilon}(U, M) \colon \mathbb{R}^3 \to \{-1, 1\}$$

as

$$s_{r,\epsilon}(U, M)(x) := \begin{cases} 1 \text{ if } U \text{ is over-segmented at } x \text{ compared to } M \\ -1 \text{ if } U \text{ is under-segmented at } x \text{ compared to } M. \end{cases}$$

In the discretized implementation, images are represented by voxels where each voxel has an associated center point denoted by x_j. $B_r(x_j)$ is given by a $3 \times 3 \times 3$ voxel neighborhood around x_j and denoted by $N(j)$. Choose $\epsilon > 0$ so that it equals the volume of a single voxel. Over-segmentation at a voxel is detected when the current voxel of the segmented region U has a 1 while the model M has a 0 and at least one neighbor of the current voxel in the segmented region has a 1. Over segmentation generally occurs because of insufficient gradient at a nearby feature boundary, often caused by blurring or distortion of features that is typical of limited-angle tomography reconstruction. Under- segmentation is detected when the current voxel of the segmented region has a 0 while the model has a 1 and at least one neighbor of the current voxel in the segmented region has a 1. Under segmentation occurs due to low intensity value of the voxel. Figs. 4.1 and 4.2 illustrates over- and under-segmentation in the 2D case.



1	0	0
1	$\chi_M(x_j) = 0$	0
0	0	0

Model M

1	1	0
1	$\chi_U(x_j) = 1$	0
1	0	0

Over segmentation of U at x_j

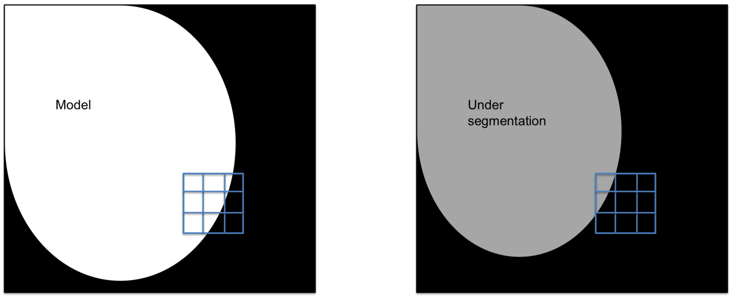
Figure 4.1: A pixel to pixel comparison between the model and segmented reconstruction over a given neighborhood detects over-segmentation at jth pixel x_j. The two tables below show the values of the indicator function for a neighborhood region of the model and corresponding segmented region.

4.2.2 Energy functional for local segmentation information

Given a model M and a segmentation region U (both domains of $\mathbb{R}^3$), the energy functional is defined as:

$$S(f, U) := \int_{\mathbb{R}^3} \Big(1 + \gamma_{r,\epsilon}(U, M)(x)\Big) ||\nabla f(x)|| dx \tag{4.8}$$

where $\gamma_{r,\epsilon}(U, M)(x) := s_{r,\epsilon}(U, M)(x) \triangle_r (U, M)(x)$.



1	1	0
1	$\chi_M(x_j) = 1$	0
1	1	0

Model M

1	0	0
0	$\chi_U(x_j) = 0$	0
0	0	0

Under segmentation of U at x_j

Figure 4.2: A pixel to pixel comparison between the model and segmented reconstruction over a given neighborhood $\mathcal{N}(j)$ detects under-segmentation at jth pixel x_j. The two tables show the values of the indicator functions for a neighborhood region of the model and corresponding segmented region.

4.2.3 Reconstruction scheme

The reconstruction method is an intertwined iterative scheme given as:

$$\begin{cases} U^{(i)} \subset \mathbb{R}^3 \text{ is segmented based on } f^{(i)} \\ f^{(i+1)} := \arg\min_f \lambda S(f, U^{(i)}) + D\big(H(f), g\big). \end{cases} \tag{4.9}$$

The segmentation method used to get $U^{(i)}$ from $f^{(i)}$ is described in Section 4.2.4. This is an iterated regularization method (i.e. solving a sequence of regularizations) where the iterate updating f provides the MAP estimate using the prior probability distribution in (4.3) with S as in (4.8). The segmentation region $U^{(i)}$ is updated while the model M is fixed. When convergent, the set $U^{(i)}$ will not change and the entire iterative sequence will tend to a MAP estimator.

The inner iterative scheme based on (4.6) is discretized to calculate the

MAP estimate $f^{(i+1)}$. The reconstruction update at each voxel is given as:

$$f_j^{(i+1)} := \frac{f_j^{(i)}}{1 + \lambda |\partial_j S_j(f_j^{(i)}, U^{(i)})|} \left[H^T \frac{g}{Hf^{(i)}} \right]_j \tag{4.10}$$

where

$$S_j(f_j, U) := \Big(1 + \gamma_j(U, M)\Big) \sum_{k \in N(j)} \frac{|f_j - f_k|}{d_{j,k}} \tag{4.11}$$

is the resulting energy functional at voxel j. Here, $d_{j,k}$ denotes the Euclidian distance between the midpoints x_j and x_k of the jth and kth voxels, and $\gamma_j(U, M)$ is the discretization of $\gamma_{r,\epsilon}(U, M)(x_j)$ where the continuum neighborhood $B_r(x_j)$ is replaced by the voxel neighborhood $N(j)$ and ϵ is chosen so that it corresponds to the volume of one voxel. The term $\triangle_r(U, M)(x)$ depends on local neighborhood matches between U and M; more matches imply higher confidence. Depending on whether over or under segmentation is detected, the term $s_{r,\epsilon}(U, M)(x)$ controls the effect of $\gamma_{r,\epsilon}(U, M)(x_j)$. This attempts to reduce blurring or distortion by controlling the effect of the regularization locally in the reconstruction.

The penalized EM optimization method was found to be superior when compared to a Conjugate Gradient optimization approach; in that good reconstructions were obtained in fewer iterations. For a stopping criterion the root mean square (RMS) difference between the current reconstruction and the previous iteration reconstruction, normalized over an intensity range between 0 and 1, was calculated. If the overall RMS difference between iterations is lower than 0.07 for two consecutive iterations, the iterations were terminated.

The reconstruction was executed for at least 10 iterations before this check was performed.

4.2.4 Segmentation

The segmentation method described in [50] was used to segment a region of interest U from the reconstructed 3D volume f. The method is a variant of the fast marching algorithm developed by Sethian [43]. The level set evolution of the advancing wave front is given by:

$$\phi_t + \mathbf{F}|\nabla\phi| = \mathbf{0} \tag{4.12}$$

ϕ is the level set function and $\mathbf{F}$ is a numerical force field. Two point classes are seeded on f, one for the foreground feature and another for the background. These seeds are treated as initial snakes or contours whose movements are directed by an external force field determined by the normalized gradient vector field [50], [2]. When the snakes stop evolving, a segmented volume U is obtained. Using an initial back projection reconstruction, without any regularizers, the seed points for features corresponding to the known model are provided by the user. These points are used at every iteration of the reconstruction to obtain a new segmentation. This algorithm is adequate for the objective of segmenting features like virus spike regions robustly with acceptable variability. Similar segmentations were also obtained through a combination of thresholding and connected component analysis. Since the segmentation is performed on small local regions of features, the reconstruction process was fairly invariant to the choice of segmentation technique.

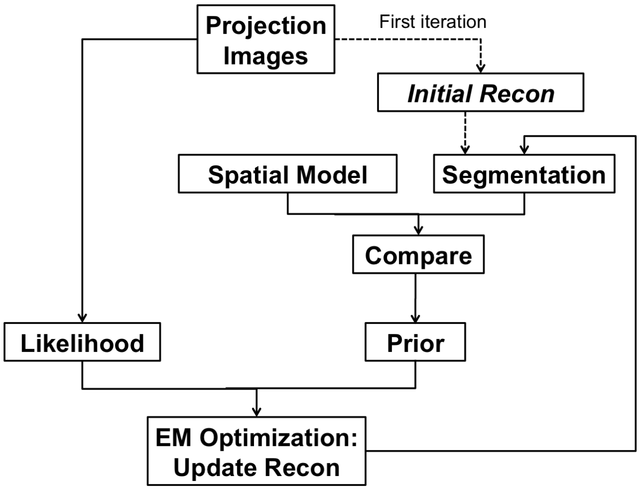


Figure 4.3: Flow diagram of the reconstruction algorithm. An initial reconstruction was used to locate regions having features that match a known spatial model. Regularization was performed in this region by performing segmentation of the updated reconstruction at each iteration and comparing the segmented region with the known spatial model.

4.3 Biological spatial models

Spatial models of the AIDS virus spike were used in the regularization step. The spike is a mushroom-shaped structure with tapering stem that is attached to the outer envelope of the virus. Using X-ray crystallography, biochemists have been able to obtain accurate atomic level information of the spike proteins of the AIDS virus. EM based volumetric models of the spikes have been generated by averaging several thousand images of the virus spike [25] (Fig. 4.4). This data is available in the Protein Data Bank (PDB) and the EM Data Bank (EMDB). The spike is the primary target for drug design as it is responsible for infecting the host cell. Accurate reconstructions of the regions surrounding the spikes can help biologists to better understand the

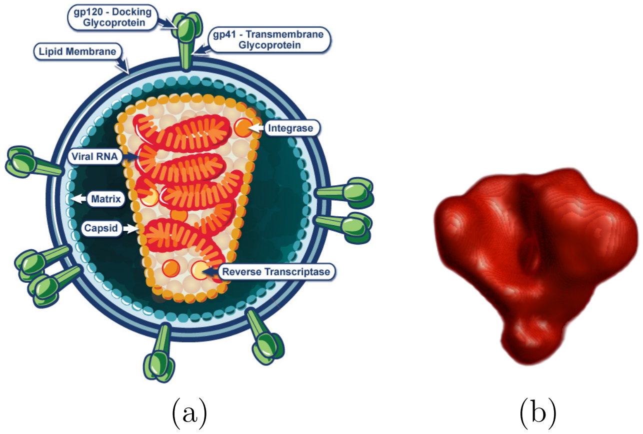


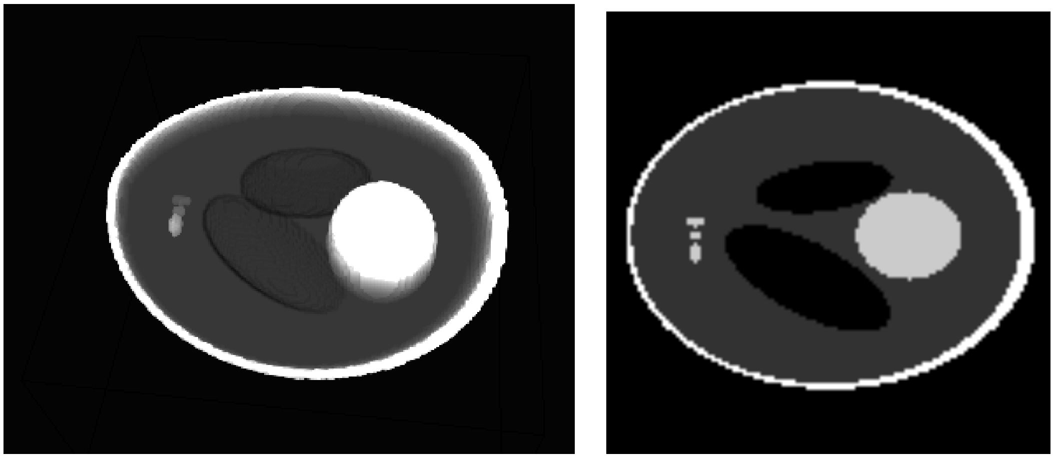
Figure 4.4: (a) A schematic of HIV AIDS virus. The mushroom shaped virus spikes are the gp120 and the gp41 regions that protrude out of the virus envelope. (source: www.niaid.nih.gov) (b) Surface Rendering of the AIDS virus spike in un-bound state [25]. The virus spike is used as the model M in the regularization process.

process with which the virus infects the host cells.

The spike model is aligned to match the spikes visible in the initial tomographic reconstruction volume (a back projection reconstruction, without any regularizers). The alignment, currently done through user supervision, needs to be done only once. The regularization is performed over subsequent iterations in a sub-volume region containing the spike and compared with the aligned model.

4.4 Results and Evaluation

This reconstruction method was evaluated on projection simulations from a 3D Shepp Logan digital phantom, a TEM simulator and experimental tilt series data from a TEM. The shape based regularization (Shape Reg)



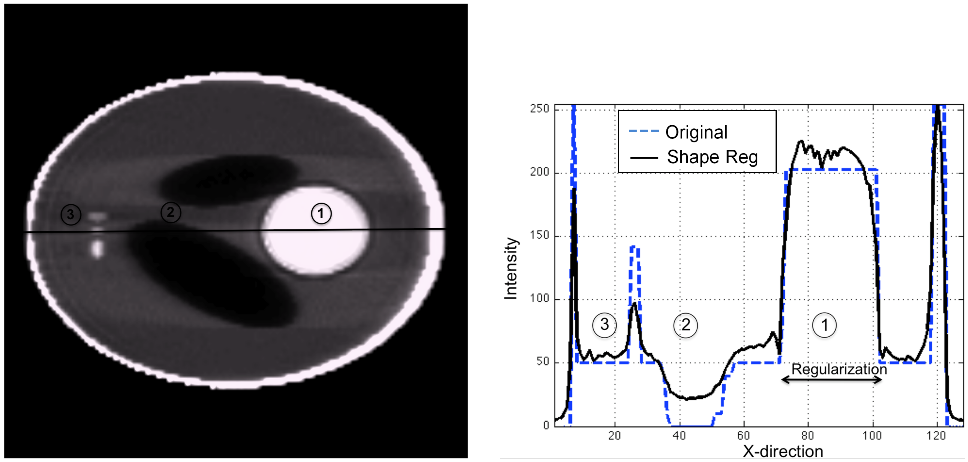
3D Shepp Logan phantom   2D Slice Shepp Logan phantom

Figure 4.5: Original 3D Shepp Logan Phantom: Volume rendering with a clipping plane of the original 3D Shepp Logan phantom is shown along with a 2D slice of the phantom. Projections of this phantom were simulated reconstruction was performed using Shape Reg and compared with reconstructions from ML-EM, WBP, L2GF and ART methods.

reconstruction output was compared with reconstructions from four different methods, WBP, ART, a L^2 gradient flow based classical regularization method (L2GF) (described in Section 1.3 [49]) and a Maximum Likelihood approach, which is the current EM implementation without the prior term (ML-EM).

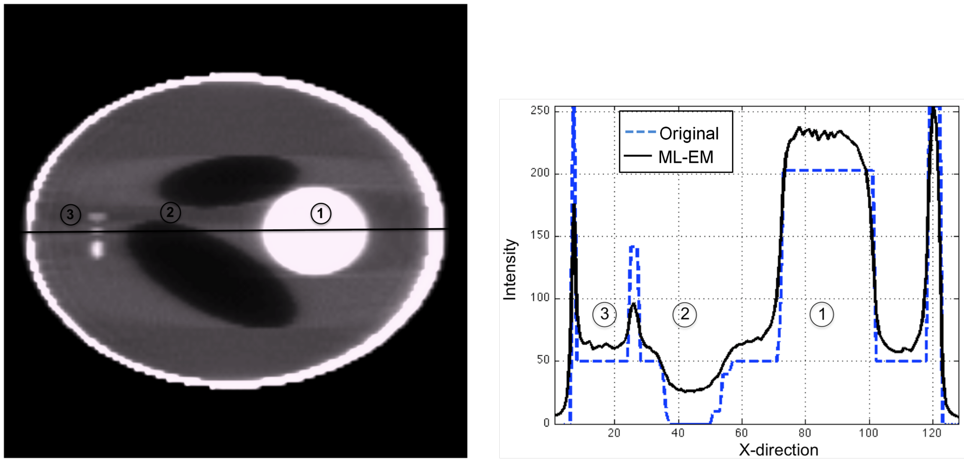
4.4.1 Digital phantom

2D projection images of a $128 \times 128 \times 128$ 3D Shepp Logan phantom (Fig. 4.6) were simulated between tilt angles $\pm 69^\circ$ in steps of 2°. The spherical region (region ① in Fig. 4.6) in the phantom was used as the known spatial model M in the Shape Reg reconstruction method. 2D central cross sections were used to compare Shape Reg with WBP, ART, L2GF and ML-EM reconstructions (Fig. 4.6).



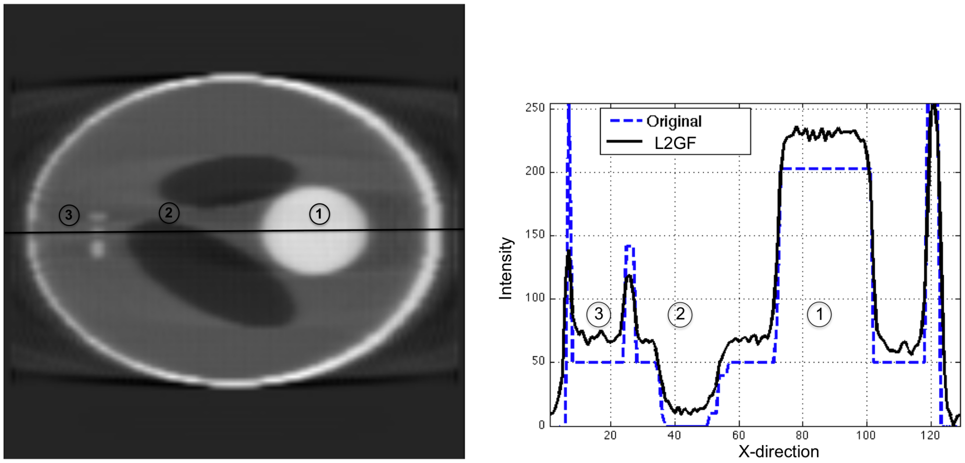
Shape Reg

Shape Reg 1D intensity profile



ML-EM

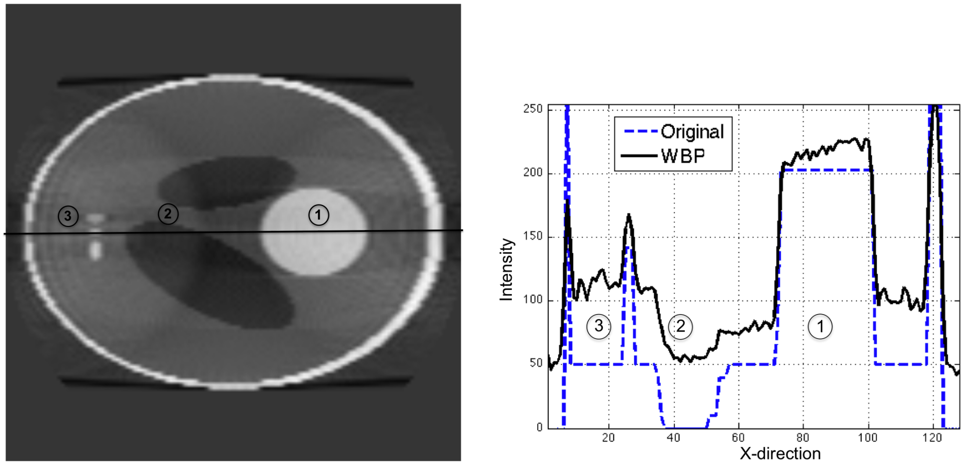
ML-EM 1D intensity profile



L2GF

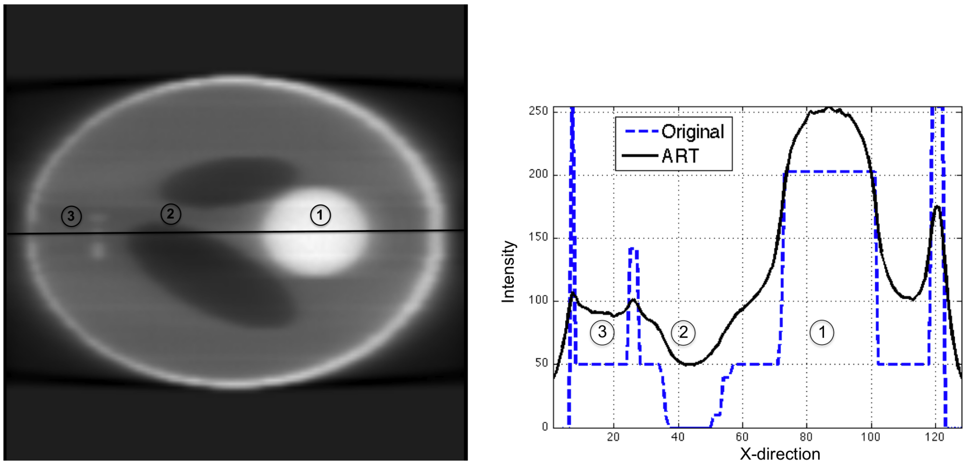
L2GF 1D intensity profile

Figure 4.6



WBP

WBP 1D intensity profile



ART

ART 1D intensity profile

Figure 4.6: Reconstruction comparison of Shepp Logan phantom: 2D central slices of Shape Reg, ML-EM, L2GF, WBP and ART reconstructions and 1D intensity profiles corresponding to the line drawn in the 2D image is shown along with the original Shepp Logan phantom's 1D profile. Regularization was applied at region ① in the Shape Reg reconstruction. WBP has sharp edges but suffers from high noise and ringing artifacts while ART appears blurred. Shape Reg performs best especially in the region ① where regularization is provided. In region ③ and the adjoining peaks Shape Reg performs best. Though no regularization was provided in regions ② and ③, Shape Reg performs better than ML-EM.

4.4.1.1 Spatial error

1D intensity profiles of the reconstructions are plotted in Fig. 4.6. This 1D profile is across the Y axis through the central X and Z coordinate as illustrated in the line through the corresponding 2D central slice shown in Fig. 4.6. Three regions of interest are marked as ①, ② and ③. Regularization for the Shape Reg method was provided for region ①. Shape Reg reconstruction intensity profile shows the best match to the profile of the phantom. In the region ① where the regularization was applied, the Shape Reg expectedly performs well (Fig. 4.6). It shows sharp edges without loss of structure fidelity. WBP's intensity profile suffers from noise and artifacts in all the regions. ART and ML-EM show a higher intensity curve that smooths the edges. L2GF performs best at region ② but both L2GF and ART show reduced intensity in the first peak in region ③. It is interesting to note that regions where regularization was *not* applied (region ③ and region ② fig 4.6) also show improved reconstruction in Shape Reg when compared with ML-EM. This implies that providing a local regularization has a global effect. The optimization step has a forward projection Hf and a back-projection H^T operator. Corrections that are made in local regions of f will affect the entire volume f due to these operators.

Mean Squared Error (MSE) was measured through a central region (2048 voxels) in each of the reconstructed volumes. The intensity range for each of the reconstruction volumes was normalized to $0 - 255$. The error for Shape Reg was the lowest at 489.7, while that of L2GF and ML-EM were close at 905.8 and 746 respectively. Both WBP and ART showed MSE values of 2302 and 3150.

4.4.1.2 Fourier spectrum plots

To analyze the effect in frequency (resolution) space, the power at various resolutions was compared. The reconstructed volumes were Fourier transformed in 3D, and the radial power spectrum was calculated. The power spectrum for the Shape Reg was compared with that of the original Shepp Logan phantom, see Fig. 4.7. All methods showed some amount of blurring compared to the original spectrum, but the Shape Reg's Fourier spectrum best matched the original Shepp Logan's spectrum. L2GF and ART exhibited the greatest amount of blurring.

4.4.2 Electron Tomography Simulation

A TEM simulator [40] was used to simulate tilt series projections of a group RNA polymerase particles. Projections were simulated with and without TEM noise between ±60° in steps of 2°. Shape Reg reconstruction was performed and the output volume was compared with the original RNA polymerase phantom volume. Reconstruction was also performed with a misaligned prior model to evaluate the robustness of the Shape Reg method.

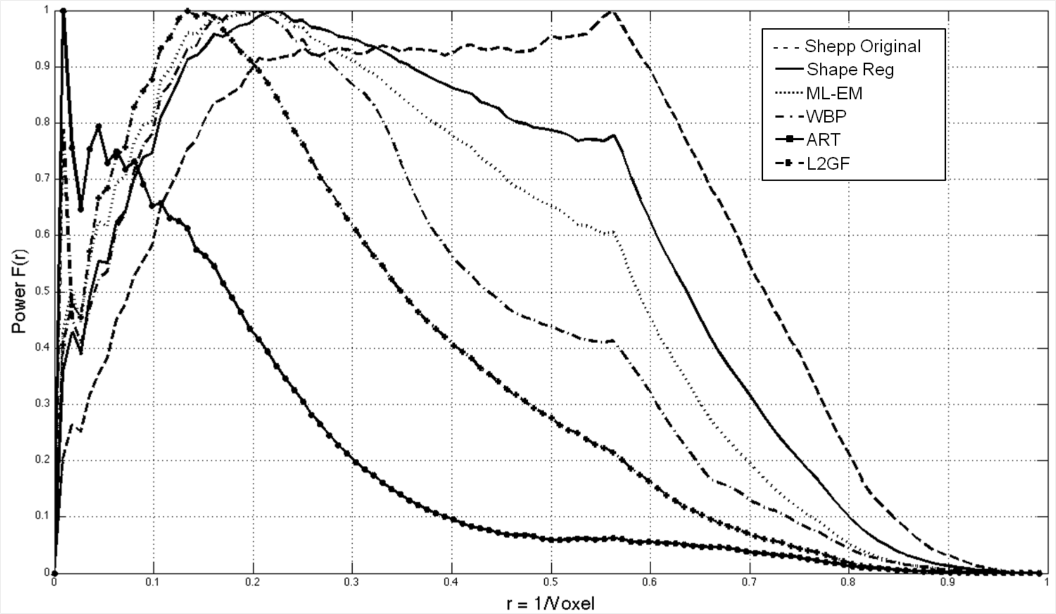
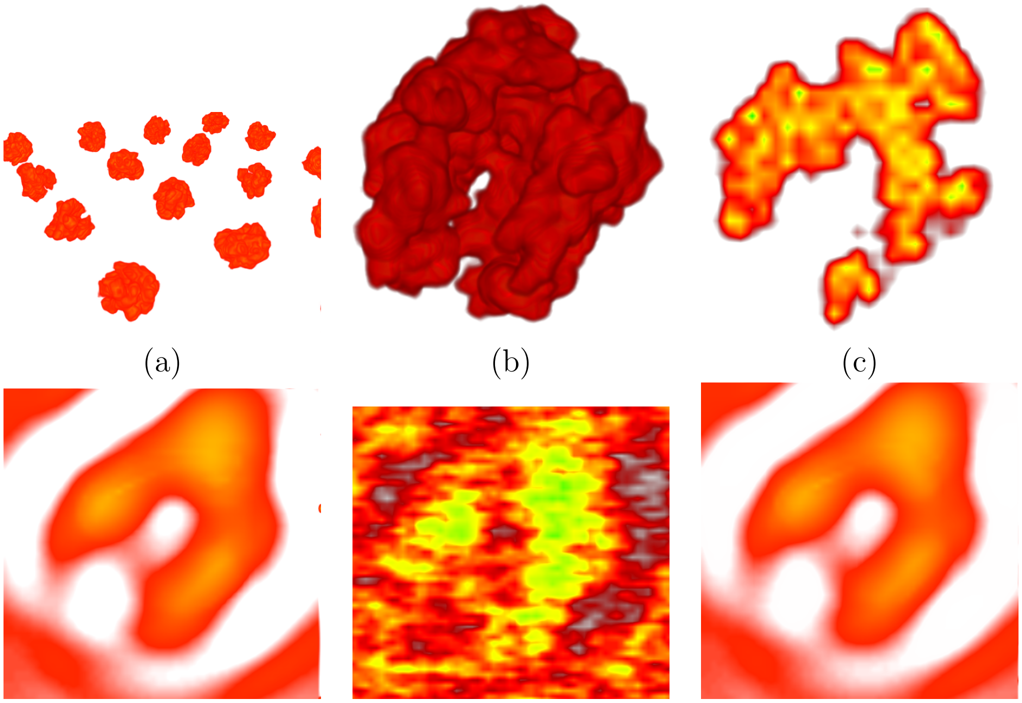


Figure 4.7: Radial Fourier power spectrum plot. The spectrum of Shape Reg is closest to that of the original phantom's spectrum. At all spatial frequencies the Shape Reg spectrum closely follows that of the original.



(a) (b) (c)

(d) (e) (f)

Figure 4.8: RNA polymerase phantom for TEM simulation (a) 3D Volume rendering (VR) of all the particles, (b) 3D VR of the central RNA particle, (c) 2D slice of RNA particle. Reconstruction of TEM simulations: 2D slice view of Shape Reg reconstruction of central RNA particle (d) with no TEM noise (e) with TEM noise and (f) misaligned prior model without TEM noise.

4.4.2.1 Without Noise

Reconstructions were performed without TEM noise using the Shape Reg method, where a part of the original RNA polymerase was used as a regularizer. The reconstructions were compared with the original phantom and the mean squared error (MSE) measured for the central RNA polymerase particle was 695.61.

4.4.2.2 With Noise

The Shape Reg reconstructions performed above were repeated with TEM noise in the tilt series projection simulations. The MSE for the reconstructed volume was 2101.77. The images were severely corrupted with noise but the structure was still discernible (Fig. 4.8(e)).

4.4.2.3 Rotated Spatial Model

To evaluate the robustness of the Shape Reg method to misalignments, the prior model was rotated by 15° about its center. The Shape Reg reconstruction was then compared with the Shape Reg reconstruction obtained previously without misalignment. Visually the Shape Reg reconstruction with misalignment appears good with no artifacts that can be attributed to the misalignment. The MSE for Shape Reg with misalignment was only slightly higher at 704.15 compared with the MSE for Shape Reg without misalignment at 695.61. This indicates that Shape Reg reconstruction is robust under small misalignments.

4.4.3 Electron tomography data

Tilt series projection images of the Simian immunodeficiency virus (similar to the HIV, AIDS virus) particles interacting with D1D2IgP neutralizing drug were acquired between angular tilts ±69° [3] using a TEM. Shape Reg reconstruction was performed using spatial models of the spike region as regularizers (see Section 4.3). The model was applied to eight spike regions whose locations were identified by a knowledgeable user in an initial reconstruction performed by a simple back-projection. The resulting Shape Reg reconstruction, shown in Figs. 4.9 and 4.10, was compared with WBP, ART, L2GF and reconstruction with no regularization (ML-EM). The spikes looked visually more distinct with better spatial differentiation between neighboring spikes. Blurring, a typical ailment of limited angle tomography, was substantially reduced in the Shape Reg method. In the regularization process, a segmentation step was used where blur in these regions resulted in over segmentation that was detected and corrected.

The neighboring regions of the spike (D1D2IgP in Fig. 4.9) also showed an improvement though no regularization was explicitly provided here. A spike region in a neighboring virus also showed improvement in its overall structure when compared to WBP, ART, ML-EM and L2GF. As seen in the digital phantom experiment, regularization at one region improved the reconstruction in nearby regions.

4.4.3.1 Resolution through Fourier Shell Correlation (FSC)

The FSC method is a common approach to estimate the resolution and quality of reconstructed volumes in ET [9] [47]. To calculate FSC, the projection images were split into two sets of alternating projection images. Reconstruction was performed on the two sets independently. The resulting reconstruction outputs were then correlated in Fourier space. The spatial frequency at which the correlation value drops to 0.5 of its maximum was estimated as the resolution (Fig. 4.11). The resolution for ML-EM, ART, WBP, and L2GF was measured to be around 31Å, 28Å, 30Å and 28Å respectively, while the Shape Reg method measured at 22Å.

4.4.3.2 Contrast to Noise Ratio (CNR)

The CNR was measured at various locations (Table 4.1) of the reconstruction volumes.

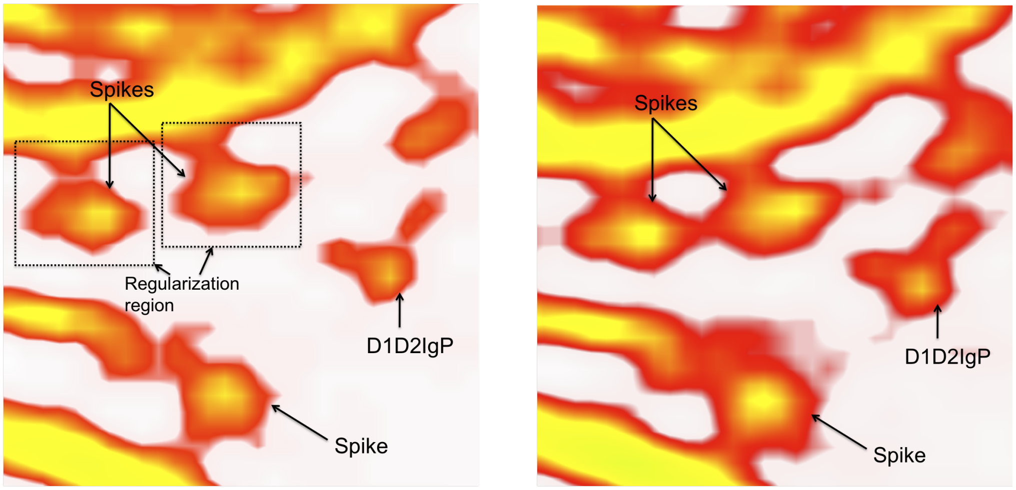
$$\text{CNR} = \frac{\mu_{\text{foreground}} - \mu_{\text{background}}}{\sigma_{\text{background}}}, \tag{4.13}$$

where $\mu_{\text{foreground}}$ is the average intensity value on a feature and $\mu_{\text{background}}$ is the average intensity value in the adjoining background region. $\sigma_{\text{background}}$ is the standard deviation measured in the background region. The CNR was measured at features like virus envelope region and spikes including those on which Shape based regularization was applied. Overall, Shape Reg gave higher CNR values. Locations $5, 6, 7$ (Table 4.1) are spike regions where regularization was applied in Shape Reg reconstruction. The regularization decreased blurring in Shape Reg, especially between neighboring spikes leading to sig-

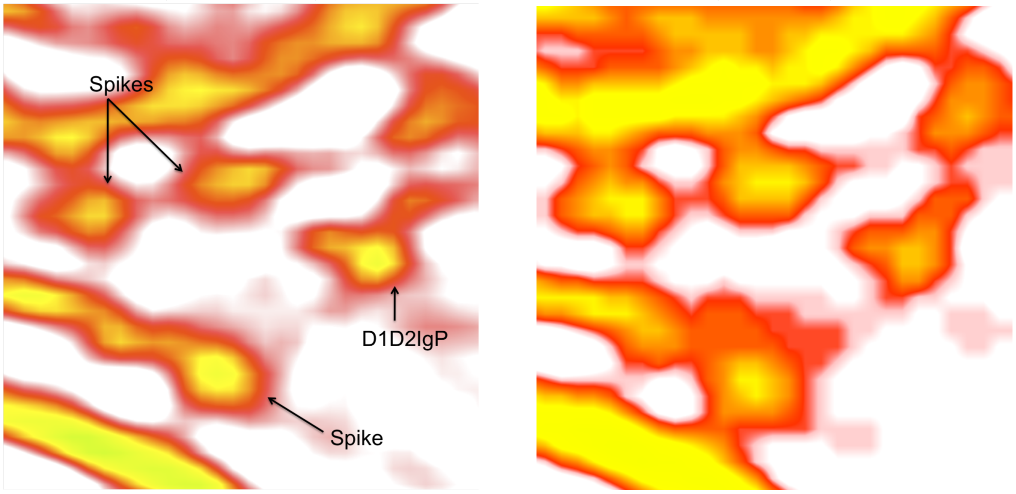
Location	Shape Reg	ML-EM	WBP	L2GF	ART
1	5.8	5.3	4.6	4.7	4.9
2	5.1	4.2	2.1	3.6	3.1
3	4.0	3.3	2.5	3.7	2.8
4	11.2	9.8	6.7	7.3	4.2
5	16.1	9.9	4.7	5.4	6.6
6	17.7	10.4	5.2	5.6	7.3
7	18.5	13.8	5.4	3.5	5.7
8	11.6	10.0	4.6	6.9	6.3
9	10.0	7.8	8.8	8.7	6.8
10	5.0	4.2	3.3	7.4	3.9
μ (p)	10.5	7.8 (0.002)	4.8 (0.002)	5.7 (0.0195)	5.1 (0.002)

Table 4.1: CNR values computed at different locations of the reconstructed volumes. The average CNR value (μ) and the p value computed using Wilcoxon paired T test between Shape Reg and other reconstruction method's CNR is shown.

nificantly higher CNR values. Other methods blur the region between spikes resulting in decreased CNR. A paired difference test of the CNR values was performed and the p value was computed. Since the CNR values follows a non-normal distribution, Wilcoxon T tests (also called signed-rank test) were performed between CNR values obtained for Shape Reg and other reconstructions method's CNR values. $p < 0.02$ was obtained for all the comparisons indicating that the CNR improvement seen in Shape Reg is statistically significant.



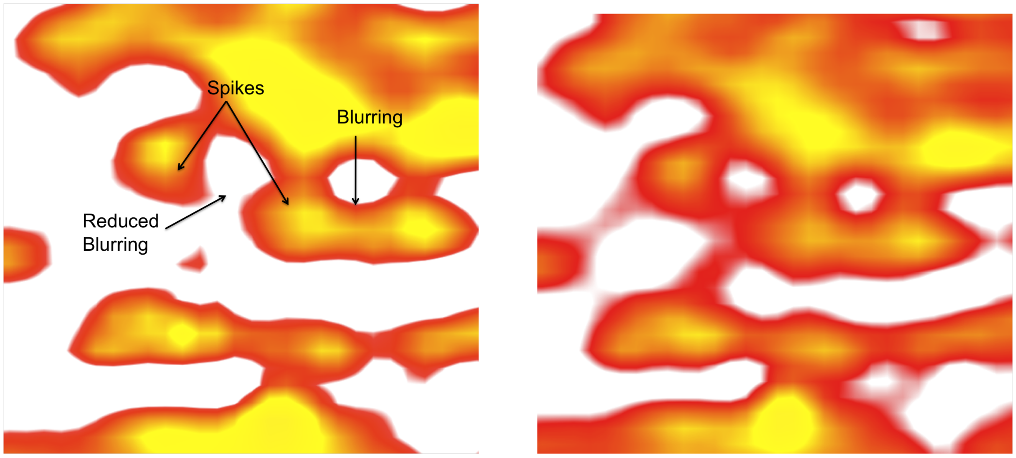
(Shape Reg) (ML-EM)



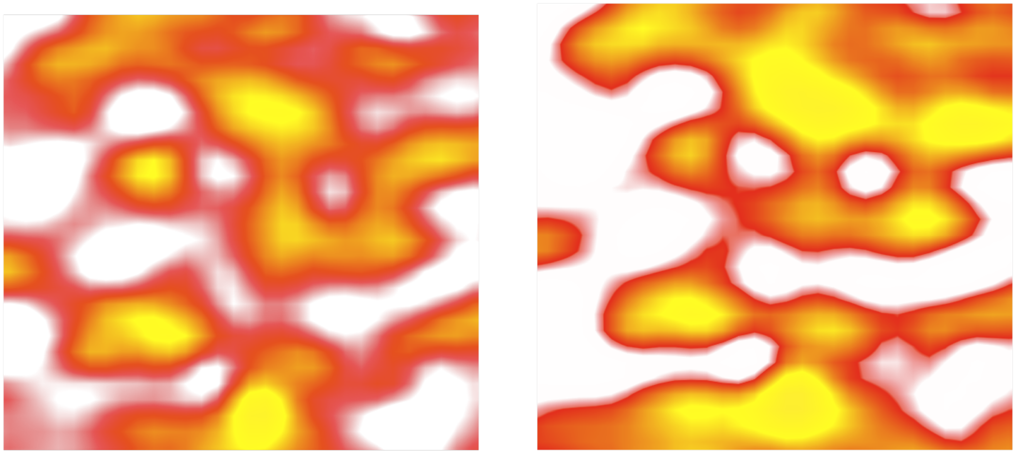
(WBP) (ART)

(L2GF)

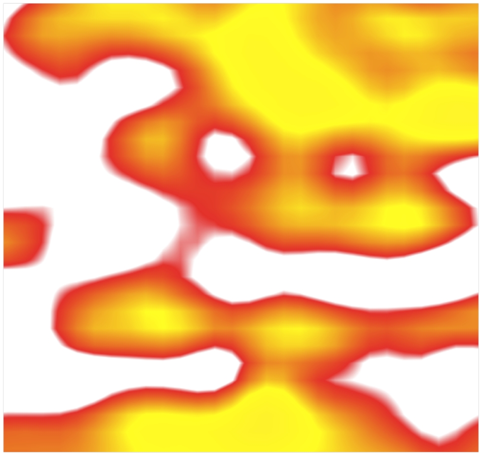
Figure 4.9: 2D slices of the reconstructed volumes showing the virus spike regions. Regularization (Shape Reg) through a prior model was applied *only* to the two spike regions shown. The regions between spikes show blurring in the other reconstruction methods. The Shape Reg method shows reduced blurring overall. An improvement can also be seen in the neighboring regions where no regularization was applied.



(Shape Reg) (ML-EM)



(WBP) (ART)



(L2GF)

Figure 4.10: 2D slices of reconstructed volumes. Regularization (Shape Reg) was applied to three spike regions. Blurring between two of the spikes was severe and there was no reduction even with regularization. However, blurring was substantially reduced around one of the spikes. In other reconstruction methods, the regions between all the spikes show severe blurring and the spikes are not distinct.

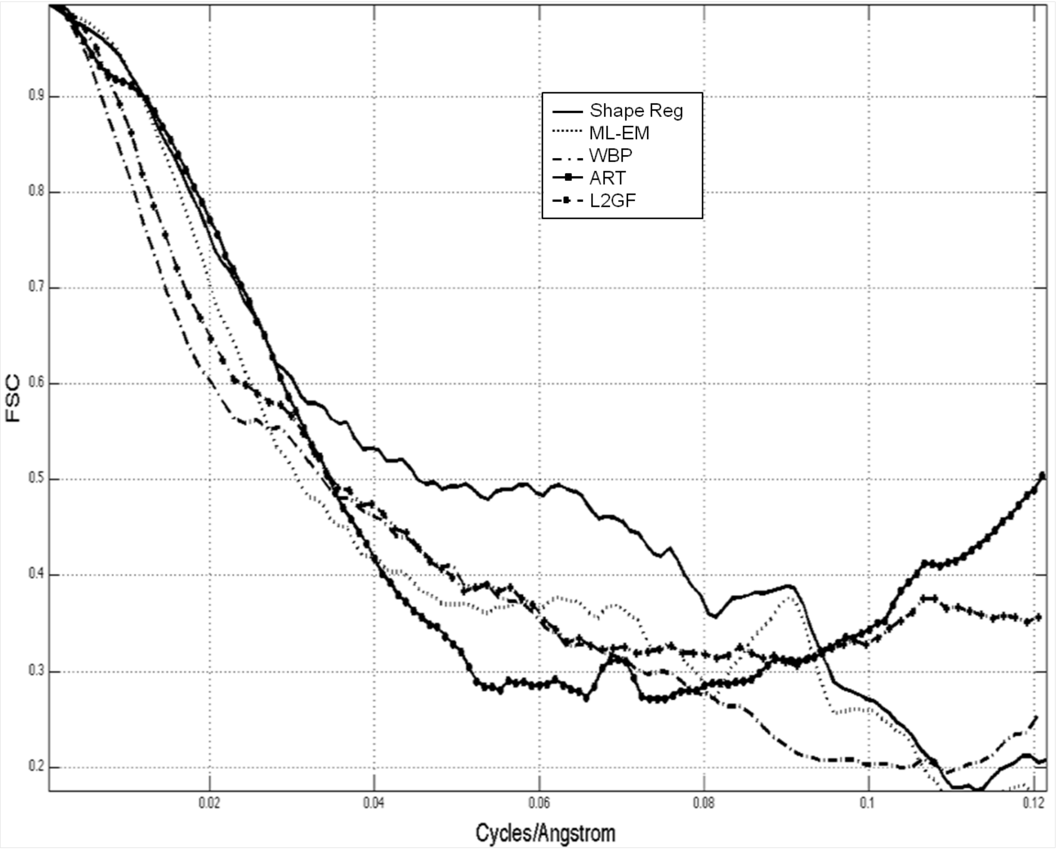


Figure 4.11: Measuring resolution through Fourier Shell Correlation (FSC): The 0.5 FSC drop off for ML-EM was 31Å, for WBP was 30Å, ART and L2GF were at 28Å while Shape Reg showed an improved resolution of 22Å

Chapter 5

Conclusion

A fully automated technique to extract the spike features of the AIDS virus has been introduced. This method uses biological and structural information about the AIDS virus and the spike position and orientation vis-a-vis the virus to detect and extract these spikes. 3D volumetric images of the AIDS virus reconstructed from tilt series projection images generated from an electron microscope was used.

This method is the first fully automated technique that can extract subvolumes of the AIDS virus spike and build a statistical model without the need for any manual supervision. This is a significant improvement over current methods (Liu et al. [25], Zhu, et al. [54]) where biologists and biochemists use manual supervision to extract spikes and build a single average model. This method can accelerate and increase the image data processing capacity of biochemists who seek to build models of the AIDS virus. Increased sample size as a result of larger data processing can lead to more accurate models of the virus spike. Shape complementarity between the spike and drug molecule is critical for the drug to effectively bind with the spike and neutralize the virus. Powerful statistical shape models can help in better drug design strategies.

Currently this statistical model is based on 72 individual spike samples. With access to a larger set of electron microscopy data of the AIDS virus, more powerful statistical models can be built. Higher resolution data would also help, as the FP rate would be reduced leading to more ideal FROC curves. Using the tools developed for this method, one can analyze and build models of the AIDS virus envelope and other features of interest. Through minor modifications, this method can be easily extended to detect structures on the envelope of other virus and bacteria particles, which would be of interest for drug design.

Secondly, a framework for including known spatial models of critical small structures in a tomographic reconstruction process has been presented. This is done through a regularization step where spatial models are introduced through a prior term. This technique does not impose the spatial model directly but influences the shape of the features by varying the prior function through an intermediate segmentation step. Specific regions with known spatial structures are segmented at each iteration and compared to a model. Segmentation errors, i.e. under and over segmentation, are corrected by appropriate variations in the regularizer. Typical regularization approaches assume a global prior characteristic like smoothness, which may not be valid in regions with very small features. This regularization, driven locally by determining local shape information through segmentation, ensures more accurate reconstruction and reduces blurring that is typical in limited angle tomography. The local regularizers also help improve the reconstruction in neighboring regions that do not have regularization applied as seen in the Shepp Logan phantom experiments, and in the improved resolution of the virus data. The application of local regularizers appears to provide a global improvement of spatial resolution, as illustrated by the Fourier power spectrum and confirmed by Fourier Shell Correlation analysis. This method also yields statistically significant improvement in contrast to noise ratios when compared with other reconstruction methods. This method also leads to improved visualization of features, which is important in the context of sub-volume averaging [51]. Methods such as WBP, which are commonly used for reconstruction in ET, can lead to erroneous selection of sub-volume features causing poor or even flawed results. Accurate reconstructions obtained by this method can further aid in solving biological issues involving drug discovery and design.

This method can further be useful in the reconstruction of tilt series of stained tissue samples that utilize small gold beads to facilitate alignment. The spherical shape of these beads can provide a model to facilitate global improvement of the reconstruction. While imaging a biological complex, neighboring structures or certain substructures whose models are known can be used as local regularizers. This will reduce the blurring effects in those local regions and improve the overall reconstruction of the biological complex. This approach can be further improved by incorporating probabilistic shapes and flexible fitting models as priors.

Appendix

Appendix 1

Physics of Image Formation

An incident electron interacts with the electrostatic potential field of a specimen resulting in scattering. The scattered electron provides information about the electrostatic properties of the specimen. Various types of scattering:

- Elastic: Energy is not transferred to the specimen
- Inelastic: Energy is transferred to the specimen
- Coherent: No phase change, typically in elastic scattering
- Incoherent: Phase change, typically in inelastic scattering

Depending on the angle of scattering it is classified into forward scattering and back scattering. For images we are interested in Forward scattering. Most biological specimens are weakly scattering, hence they do not deviate too far from the incident beam. Also, scattering strength decreases at higher angles. Most of the scattered electrons are within $\pm 5^\circ$ of un-scattered beam [48].

The main contrast mechanism in TEM imaging is phase contrast (interference contrast), which results from the quantum superposition (interference) of the crests of the single electron wave as it passes through the specimen. (Note: Wave nature of electrons and photons are different) The optical system of the microscope converts the phase contrast to visible amplitude contrast. The measured intensity for a single electron is the quantum superposition of a set of intensities each belonging to a coherent subspace. (Coherent subspace: ex: inelastic and elastic scattering are different coherent subspaces) [8].

1.1 Born Approximation

Let $u = e^{i\vec{k}.\vec{x}}$ be an incoming wave [22] whose wave number (spatial frequency) is given by $k = \frac{\lambda}{2\pi}$ at position $\vec{x}$. For homogeneous media the wave number is constant. The wave equation (Helmholtz equation) is given by

$$(\nabla^2 + k^2)u(\vec{x}) = 0 \tag{1.1}$$

Let $u^{out} = u^{sc} + u^{in}$ where u^{out} is the outgoing wave and u^{sc} is the scattered wave and u^{in} is the incident wave. Assuming the Sommerfeld radiation condition according to which energy propagates from the source and scatters to infinity and no energy can be radiated into the source. u^{out} satisfies the Helmholtz equation:

$$(\nabla^2 + k^2)u^{out}(\vec{x}) \;\; = \;\; -f(\vec{x})u^{out}(\vec{x}) \tag{1.2}$$

Where $f(\vec{x})$ is the scattering function (potential) due to inhomogeneity in the medium. In electron microscopy $f(\vec{x}) = \frac{-2m}{h^2}V(x)$

substituting for $u^{out} = u^{sc} + u^{in}$,

$$(\nabla^2 + k^2)(u^{sc}(\vec{x}) + u^{in}(\vec{x})) = -f(\vec{x})(u^{sc}(\vec{x}) + u^{in}(\vec{x})) \tag{1.3}$$

Since u^{in} satisfies equation 1.1,

$$(\nabla^2 + k^2)u^{sc}(\vec{x}) = -f(\vec{x})(u^{sc}(\vec{x}) + u^{in}(\vec{x})) \tag{1.4}$$

The solution to the above equation 1.4 is written in terms of a Green's function G_k which is a solution to the differential equation [8] [22] [29]

$$G_k(\vec{x} - \vec{x'}) = \frac{e^{ik(\vec{x}-\vec{x'})}}{4\pi(\vec{x} - \vec{x'})}$$

$$(\nabla^2 + k^2)G_k(\vec{x} - \vec{x'}) = \delta(\vec{x} - \vec{x'}) \tag{1.5}$$

Multiply both sides of equation 1.5 by RHS of equation 1.4 and integrate over $\vec{x'}$

$$\int (\nabla^2 + k^2)G_k(\vec{x} - \vec{x'})(-f(\vec{x'})(u^{sc}(\vec{x'}) + u^{in}(\vec{x'})))d\vec{x'} = \\ \int \delta(\vec{x} - \vec{x'})(-f(\vec{x'})(u^{sc}(\vec{x'}) + u^{in}(\vec{x'})))d\vec{x'}$$

rearranging the terms and evaluating the integral on RHS, which is similar to the sifting property for a function

$$(\nabla^2 + k^2)\int G_k(\vec{x}-\vec{x'})(-f(\vec{x'})(u^{sc}(\vec{x'})+u^{in}(\vec{x'})))d\vec{x'} = -f(\vec{x})(u^{sc}(\vec{x})+u^{in}(\vec{x})) \tag{1.6}$$

Comparing equation 1.6 and 1.4,

$$u^{sc}(\vec{x}) = \int G_k(\vec{x}-\vec{x'})(-f(\vec{x'})(u^{sc}(\vec{x'})+u^{in}(\vec{x'})))d\vec{x'} \tag{1.7}$$

which is the solution for $u^{sc}(\vec{x})$. This integral equation 1.7 cannot be evaluated directly since u^{sc} appears both in the LHS and in the RHS of the equation.

If the scattered field u^{sc} is small compared to u^{in}, the equation 1.7 can be written as

$$u^{sc}(\vec{x}) = \int G_k(\vec{x}-\vec{x'})(-f(\vec{x'})u^{in}(\vec{x'}))d\vec{x'} \tag{1.8}$$

This is also called the first order Born approximation. The second order approximation can be found by substituting $u^{in}(\vec{x}) = u^{in}(\vec{x}) + u^{sc}(\vec{x})$ [22]

$$u^{sc}_{(2)}(\vec{x}) = \int G_k(\vec{x}-\vec{x'})(-f(\vec{x'})(u^{in}(\vec{x'})+u^{sc}_{(1)}(\vec{x'})))d\vec{x'} \tag{1.9}$$

where $u^{sc}_{(1)}$ is the first order Born approximation.

The wave u^{out} is,

$$u^{out}(\vec{x}) = u^{in}(\vec{x}) + \int G_k(\vec{x} - \vec{x'})(-f(\vec{x'})u^{in}(\vec{x'}))d\vec{x'} \tag{1.10}$$

For the integral in equation 1.10, Natterer [29] uses the limits $|\vec{x'}| < \rho$ where ρ is defined as $f(\vec{x}) = 0$ for $|x| > \rho$

1.2 Criteria

The field inside the object (inhomogeneity) is given by

$$u^{object} = Ae^{jk_0(1+f(\vec{x}))x} \tag{1.11}$$

The phase difference between the incident field and the field inside the object is "equal to the integral through the object of the change in refractive index" [22].

$$Phase\ Change = \int_{|x|<\rho} k_0 f(\vec{x})dx = 2k_0 f\rho \tag{1.12}$$

Note: The phase change equation in [22] is given as $4\pi n_\delta \frac{a}{\lambda}$ where a is the radius of the object. The function $f(\vec{x})$ used here is represented by [22] as n_δ, the refractive index deviation of the object from the background. n_δ appears as a constant in the integral w.r.t x if substituted for $f(\vec{x})$ in the equation 1.12. $f\rho$ would represent the value of the scattering function integrated over the bounds of the object.

For Born approximation to be valid this phase change must be less than

π [22]. The condition expressed by [22] reduces equation 1.12 to:

$$\rho f < \frac{\lambda}{4} \tag{1.13}$$

Fanelli et al. state the criteria as follows: "the size of smallest details in the object that need to be recovered times the contrast (which is defined as the relative size of the function describing the object w.r.t. the background) should be much less than π/k (i.e. $\rho < \frac{\pi}{k*f}$)." [8]

The condition stated by Natterer [29] is essentially the same as above.

$$k\ \rho\ sup_{|x|<\rho}|f(\vec{x})| < 2\pi c \tag{1.14}$$

Where k is the wave number defined as $k = 2\pi/\lambda$, ρ is the size of the object, $f(\vec{x})$ is the scattering function of the object which is assumed to correspond to the object's measured contrast on the image and c is a constant taken to be around 0.16.

1.3 Intensity function

The intensity generated by a single electron is given as [33]:

$$I_z(f) = \frac{1}{M^2}(1 - (2\pi)^{-2}[PSF \otimes \wp(f)]\frac{z}{M}k^{-1}) \tag{1.15}$$

Where $I_z(f)$ is the intensity due to scattering function f and z is perpendicular to the image plane, PSF is the point spread function that models the effects of the microscope optics (for incoherence, PSF is split into real and imaginary parts). M is the magnification, f is the scattering potential and k is the wave number. $\wp(f)$ is defined as the parallel beam transform, which is the projection of scattering function f. This imaging equation is derived from the outgoing wave equation $u^{out}(\vec{x})$ (equation 1.10) and its interaction with the optics of the microscope. This takes into account the point spread function of the overall imaging system including the aperture, lens and the detector.

Bibliography

[1] C. Bajaj, Guoliang Xu, and Qin Zhang. Smooth surface constructions via a higher-order level-set method. In *IEEE International Conference on Computer-Aided Design and Computer Graphics*, 2007.

[2] C. Bajaj, Z. Yu, and M. Auer. Volumetric feature extraction and visualization of tomographic molecular imaging. *Journal of Structural Biology*, 144:132–143, 2003.

[3] A. Bennett, J. Liu, D. V. Ryk, D. Bliss, J. Arthos, R. M. Henderson, and S. Subramaniam. Cryoelectron tomographic analysis of an hiv-neutralizing protein and its complex with native viral gp120. *The Journal of Biological Chemistry*, 282:27754–27759, 2007.

[4] F. L. Bookstein. *Morphometric tools for Landmark Data.* Cambridge University Press, 2003.

[5] D. P. Chakraborty and L. H. Winter. Free-response methodology: alternate analysis and a new observer-performance experiment. *Radiology*, 174:873–881, 1990.

[6] T.F. Cootes and C. J. Taylor. Combining point distribution models with shape models based on finite-element analysis. *Image and Vision Computing*, 13(5):403–409, 1995.

[7] T. K. Dey and S. Goswami. Provable surface reconstruction from noisy samples. In *20th ACM-SIAM Sympos. on Comput. Geom.*, 2004.

[8] D Fanelli and O. Ö Oktem. Electron tomography: a short overview with an emphasis on the absorption potential model for the forward problem. *Inverse Problems*, 24(013001), 2008.

[9] Joachim Frank, editor. *Electron tomography: Methods for Three-Dimensional Visualization of Structures in the Cell*, chapter 3, 10. Springer, 2006.

[10] Gene Gindi, Mindy Lee, Anand Rangarajan, and I. George Zubal. Bayesian reconstruction of functional images using anatomical information as priors. *IEEE Transactions on Medical Imaging*, 12(4):670 – 680, 1993.

[11] Ajay Gopinath. Reconstruction comparison and a composite segmentation method for electron tomography. In *IEEE International Conference on Image Processing*, pages 585 – 588, 2010.

[12] S. Goswami, T. Dey, and C. Bajaj. Identifying flat and tubular regions of a shape by unstable manifolds. In *Proc. Sympos. Geometry Processing*, pages 143–152, 2006.

[13] D. Gourion and D. Noll. The inverse problem of emission tomography. *Inverse Problems*, 18:1435–1460, 2002.

[14] Peter J. Green. On use of the EM algorithm for penalized likelihood estimation. *Journal of Royal Statistical Society*, 52(3):443–452, 1990.

[15] Ulf Grenander and Michael I. Miller. Computational anatomy: an emerging discipline. *Quarterly of Applied Mathematics*, LVI:617 – 694, 1998.

[16] Ulf Grenander and Michael I. Miller. *Pattern Theory: From Representation to Inference*. Oxford University Press, 2007.

[17] M. Hanke H. W. Engl and A. Neubauer. *Regularization of Inverse Problems*. Springer, 2000.

[18] Tobias Heimann and Hans-Peter Meinzer. Statistical shape models for 3d medical image segmentation: A review. *Medical Image Analysis*, 13:543–563, 2009.

[19] G.T. Herman. *Image Reconstructions from Projections: The Fundamentals of Computerized Tomography*, chapter 8. Computer Science and Applied Mathematics, 1980.

[20] Ing-Tsung Hsiao, Anand Rangarajan, Gene Gindi, and Gene Gindi. Bayesian image reconstruction for transmission tomography using deterministic annealing. *Journal of Electronic Imaging*, pages 7–16, 2003.

[21] J. Kaipio and E Somersalo. *Statistical and Computational Inverse Problems*, chapter 3. Springer-Verlag, 2005.

[22] A.C. Kak and M. Slaney. *Principles of Computerized Tomographic Imaging*, chapter 7. IEEE Press, 1988.

[23] Emanuel Levitan, Michael Chan, and Gabor T. Herman. Image-modeling gibbs priors. *Graphical Models and Image Processing*, 57(2):117–130, 1995.

[24] Ming Li, Guoliang Xu, Carlos O.S. Sorzano, Fei Sun, and Chandrajit L. Bajaj. Single-particle reconstruction using L^2-gradient flow. *Journal of Structural Biology*, 176(3):259–267, 2011.

[25] J. Liu, A. Bartesaghi, M. J. Borgnia, G. Sapiro, and S. Subramaniam. Molecular architecture of native hiv-1 gp120 trimers. *Nature*, 455:109–113, September 2008.

[26] A. K. Louis and P. Maass. A mollifier method for linear operator equations of the first kind. *Inverse Problems*, 6:427–440, 1990.

[27] Michael I. Miller. Computational anatomy: shape, growth, and atrophy comparison via diffeomorphisms. *NeuroImage*, 23:S19–S33, September 2004.

[28] Rajesh Narasimha, Iman Aganj, Adam E. Bennett, Mario J. Borgnia, Daniel Zabransky, Guillermo Sapiro, Steven W. McLaughlin, Jacqueline L.S. Milne, and Sriram Subramaniam. Evaluation of denoising algorithms for biological electron tomography. *Journal of Structural Biology*, 164(1):7–17, 2008.

[29] F. Natterer. An error bound for the born approximation. *Inverse Problems*, 20:447–452, 2004.

[30] Frank Natterer and Frank Wübbeling. *Mathematical Methods in Image Reconstruction*, chapter 5. SIAM, 2001.

[31] Robert Nürnberg. Distance from a point to an ellipsoid. Technical report, Imperial College London, 2006.

[32] S. M. Pizer, D. S. Fritsch, P. A. Yushkevich, V. E. Johnson, and E. L. Chaney. Segmentation, registration and measurement of shape variation via image object shape. *IEEE Transactions on Medical Imaging*, pages 851–865, 1999.

[33] E. T. Quinto and O. Ö Oktem. Local tomography in electron microscopy. *Society for Industrial and Applied Mathematics*, 68:1282–1303, 2008.

[34] M. Radermacher. *Electron tomography: Methods for Three-Dimensional Visualization of Structures in the Cell*, chapter 8. Springer, 2006.

[35] Anand Rangarajan, Ing-Tsung Hsiao, and Gene Gindi. A bayesian joint mixture framework for the integration of anatomical information in functional image reconstruction. *Journal of Mathematical Imaging and Vision*, 12:199–217, 2000.

[36] Torsten Rohlfing, Robert Brandt, Calvin R. Maurer, and Randolf Menzel. Bee brains, b-splines and computational democracy: Generating an average shape atlas. In *IEEE Workshop on Mathematical Methods in Biomedical Image Analysis*, pages 187–194, 2001.

[37] Michael G. Rossmann. Fitting atomic models into electron-microscopy maps. *Biological Crystallography*, D56:1341–1349, 2000.

[38] Daniel Rueckert, Alejandro F. Frangi, and Julia A. Schnabel. Automatic construction of 3-d statistical deformation models of the brain using non-rigid registration. *IEEE Transactions on Medical Imaging*, 22(8):1014 – 1025, 2003.

[39] H. Rullgård. A new principle for choosing regularization parameter in certain inverse problems. *arXiv:0803.3713v2 [math.NA]*, April 2008.

[40] H. Rullgård, L.-G. Öfverstedt, S. Masich, B. Daneholt, and O. Öktem. Simulation of transmission electron microscope images of biological specimens. *Journal of Microscopy*, 243(3):243–256, 2011.

[41] M. P. Sampat, Z. Wang, S. Gupta, A. C. Bovik, and M. K. Markey. Complex wavelet structural similarity: A new image similarity index. *IEEE Transactions on Image Processing*, 18(11):2385 – 2401, 2009.

[42] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen. *Variational Methods in Imaging.* Springer, 2009.

[43] J. A. Sethian. A marching level-set method for monotonically advancing fronts. In *National Academy of Sciences*, volume 93, pages 1591–1595, 1996.

[44] L. A. Shepp and Y. Vardi. Maximum likelihood reconstruction for emission tomography. *IEEE Transactions on Medical Imaging*, MI-1(2):113 – 122, 1982.

[45] Carlton A. Taft, Vinicius Barreto da Silva, and Carlos Henrique Tomich de Paula da Silva. Current topics in computer-aided drug design. *Journal of Pharmaceutical Sciences*, 97(3):1089–1098, 2008.

[46] B. Tsagaan, A. Shimizu, H. Kobatake, and K. Miyakawa. An automated segmentation method for kidney using statistical information. In *Proc. MICCAI LNCS*, volume 2488, 2002.

[47] Marin van Heel and Michael Schatz. Fourier shell correlation threshold criteria. *Journal of Structural Biology*, 151(3):250–262, 2005.

[48] David Bernard Williams and C. Barry Carter. *Transmission Electron Microscopy: A Textbook for Materials Science*. Springer, 1996.

[49] Guoliang Xu, Ming Li, Ajay Gopinath, and Chandrajit Bajaj. Computational inversion of electron tomography images using l2-gradient flows. *Journal of Computational Mathematics*, 29(5):501–525, 2011.

[50] Z. Yu and C. Bajaj. Computational approaches for automatic structural analysis of large bio-molecular complexes. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 5(4):568–582, 2007.

[51] Zhou Yu and Achilleas S. Frangakis. Classification of electron subtomograms with neural networks and its application to template-matching. *Journal of Structural Biology*, 174(3):494–504, 2011.

[52] R. Zanella, P. Boccacci, L. Zanni, and M. Bertero. Efficient gradient projection methods for edge-preserving removal of poisson noise. *Inverse Problems*, 25(045010), 2009.

[53] Z. Zhao, S. R. Aylward, and E. K. Teoh. A novel 3d partitioned active shape model for segmentation of brain mr images. In *Proc. MICCAI LNCS*, volume 3749. Springer, 2005.

[54] Ping Zhu, Jun Liu, Julian Bess Jr, Elena Chertova, Jeffrey D. Lifson, Henry Grise, Gilad A. Ofek, Kenneth A. Taylor, and Kenneth H. Roux. Distribution and three-dimensional structure of aids virus envelope spikes. *Nature*, pages 847–852, June 2006.

Vita

Ajay Gopinath was born in Bangalore, India on 24 September 1980, the son of Lakshmi Gopinath and K.S. Gopinath. He received the Bachelor of Engineering degree in Electronics and Communication from the Bangalore Institute of Technology (VTU) in 2002. He received a Master of Science degree in Electrical and Computer Engineering from Clarkson University , Potsdam, New York in 2005. He worked at the General Electric Global Research Center's Imaging Technology Lab 2005 to 2007. He started his PhD program at the University of Texas at Austin's Electrical and Computer Engineering department in 2007.

Permanent address: 34, 17th Cross, Malleshwaram,
Bangalore, India 560055

This dissertation was typeset with LaTeX[†] by the author.

[†] LaTeX is a document preparation system developed by Leslie Lamport as a special version of Donald Knuth's TeX Program.